2010

A Novel Synergistic Diagnosis Methodology for identifying Abnormalities in Wireless Capsule Endoscopy Videos

Alexandros Karargyris
Wright State University

A NOVEL SYNERGISTIC DIAGNOSIS METHODOLOGY FOR IDENTIFYING ABNORMALITIES IN WIRELESS CAPSULE ENDOSCOPY VIDEOS

A dissertation submitted in fulfillment of the
requirements for the degree of
Doctor of Philosophy

By

ALEXANDROS KARARGYRIS
MS., National Technological University of Athens, 2006

2010
Wright State University

WRIGHT STATE UNIVERSITY
SCHOOL OF GRADUATES STUDIES

March 17, 2010

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Alexandros Karargyris, ENTITLED A Novel Synergistic Diagnosis Methodology for identifying Abnormalities in Wireless Capsule Endoscopy videos BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy

Nikolaos G. Bourbakis, Ph.D.
Dissertation Director

Arthur Goshtasby, Ph.D.
Director, Computer Science & Engineering Ph.D. Program

John A. Bantle, Ph.D
Vice President Research and Graduate Studies

Committee on Final Examination

Nikolaos G. Bourbakis, Ph.D.

Marios Pouagare, Ph.D.

Soon Chung, Ph.D.

Thomas Hangartner, Ph.D.

Yong Pei, Ph.D.

ABSTRACT

Karargyris, Alexandros, Ph.D, Department of Computer Science & Engineering, Wright State University, 2010.

A Novel Synergistic Diagnosis Methodology for Identifying Abnormalities in Wireless Capsule Endoscopy Videos.

Wireless Capsule Endoscopy (WCE) is a new technology that allows medical personnel to view the gastrointestinal (GI) mucosa. It is a swallowable miniature capsule device the size of a pill that transmits thousands of screenshots of the digestive tract to a wearable receiver. When the procedure finishes the video is uploaded to a workstation for viewing. Capsule Endoscopy has been established as a tool to identify various gastrointestinal (GI) conditions, such as blood-based abnormalities, polyps, ulcers, Crohn's disease in the small intestine, where the classical endoscopy is not regularly used.

As of 2009 the market is dominated by Given Imaging Inc. capsule (PillCam SB). More than 300,000 capsules have been sold since 2001 when it was first introduced. The company provides a software package (RAPID) to view the WCE video, offering a bleeding detector feature based on red color. It provides a position estimator of the capsule inside the digestive tract. Additionally its multi-view feature gives a simultaneous view of two or four consecutive video frames in multiple windows. Finally a library of reference images (RAPID Atlas) is provided so that the user can have easy access to on-screen case images.

Although the company's software is a useful tool, the viewing of a WCE video is still a time consuming process (~ 2 hours), even for experienced gastroenterologists. In addition, the company's software has serious limitations

(35% bleeding detection) and no capability of detecting polyps or ulcers according to gastroenterologists. Therefore, the need for computer aided model-methodology with robust detection performance on various conditions (blood, polyps, ulcers, etc) is clearly obvious.

Thus, our research studies have been successfully carried out on: a) the automatic detection of malignant intestinal features like polyps, bleeding, and abnormal regions (tumors); b) finding the boundaries of the digestive organs; and c) reducing the viewing-examination time with a robust registration methodology. These studies have led to the development of the ATRC Video Toolbox (ATRC-VT).

ATRC-VT incorporates signal processing methods, color and image processing techniques, and artificial intelligence tools to detect blood-based abnormalities, polyps and ulcers in the small intestine. It is the first computer aided detection (CAD) software with multiple capabilities for WCE videos designed with a Graphics User Interface so that it is easy to use.

Table of Contents

Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

I would like to thank my advisor Dr. Nikolaos Bourbakis for giving me the opportunity to be creative and efficient. He encouraged me at difficult times and with tough challenges, but also his criticism helped me become a better professional.

I would also like to thank Dr. Marios Pouagare for providing the data sets for the study and valuable medical advice. Finally, I would like to thank the members of my committee Dr. Soon Chung, Dr. Thomas Hangartner, and Dr. Yong Pei for their important suggestions and support.

DEDICATION

To my beautiful parents, Maria and Christos, who made me the person I am.

To my older brother, Orestis, for his continuous support and love to me without limits.

Σας αγαπώ!

CHAPTER 1: INTRODUCTION

1.1 What is Wireless Capsule Endoscopy?

An estimated 19 million people in the United States suffer from diseases of the small intestine, such as obscure bleeding, Crohn's disease, chronic diarrhea, or cancer [1]. Wireless Capsule Endoscopy (WCE) is a technology that offers medical doctors (gastroenterologists) the ability to view the interior of the small intestine with a non-invasive procedure. WCE was invented by a group of researchers in Baltimore in 1989, and later on introduced by Given Imaging Incorporated as a commercial tool. Given Imaging has developed a swallowable wireless capsule PillCam™ SB2 (Fig. 1-1) that has been the only available product on the market since 2002. A competitor, Olympus Corporation, has been selling the EndoCapsule™ in Europe since 2005, and received FDA approval in September 2007. The difference between the two products is minor from operational point of view.

At one end of the PillCam capsule there is a miniaturized colour camera and an optical dome with four (4) white light LEDs. The camera can operate up to 8 hours and capture more than 50,000 images, which are transmitted wirelessly to a storage device worn by the patient. After the 8-hour procedure, the video feedback is uploaded to a workstation where the medical personnel can review and examine the endoscopic images.

Fig. 1-1. Common Capsule Endoscopy products in the market. Left PillCam distributed by Given Imaging and right EndoCapsule distributed by Olympus.

The capsule endoscopy market is rapidly emerging and currently in the US it is above $150 million as shown in Fig. 1-2. Projections show that in 2011 this market will exceed $370 million.

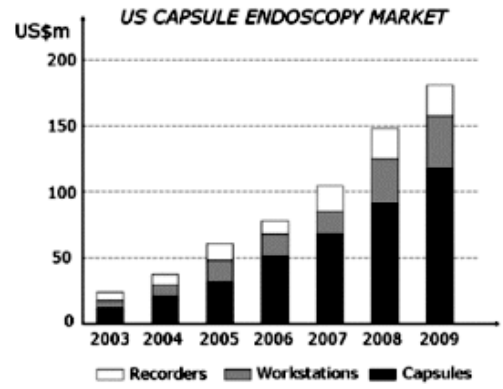


Fig. 1-2. US Capsule Endoscopy Market. Source: Millennium Research Group, Medical Industry Intelligence, January-February 2005, Volume 1, Issue 1

1.2 Wireless Capsule Endoscopy and other technologies

Besides WCE, there is a small number of other endoscopyic techniques that are used for identifying small bowel diseases.

Push Endoscopy is a procedure that allows investigation and treatment of diseases in the upper small intestine. Push endoscopy reaches further into the small intestine than the standard upper gastrointestinal endoscopy - EGD.

Sonde enteroscopy involves the use of a long, flexible, fiber optic instrument propelled through the small bowel by peristalsis; this procedure may allow for viewing the remainder of the small bowel. Sonde instruments rely on a balloon placed at the instrument's tip. Peristalsis then advances the long flexible endoscope to the distal small bowel, and the endoscopic examination is performed during withdrawal.

Intra-operative endoscopy is performed in an operating room with the patient under general anesthesia. An endoscope is inserted into the bowel by the surgeon passed through the small bowel manually while the gastroenterologist views the image. Intra-operative endoscopy is the last resort procedure to be considered in situations of severe recurrent bleeding.

Table 1-1 below synopsizes some aspects of these technologies.

Table 1-1. Comparison of Endoscopy Technologies

Feature	Push Endoscopy	Sonde Enteroscopy	Intraoperative Endoscopy	Capsule Endoscopy
Diagnostics yield for obscure bleeding	30% - 50%	26% - 77%	58% - 80%	50% - 70%
Invasive	Yes	Very	Extremely	Non-invasive
Procedure type	Impatient/Outpatient	Impatient	Impatient	Outpatient
Patient discomfort	Moderate	Severe	Extreme	Mild/none
Can view entire small intestine?	No	Yes	Yes	Yes
Therapeutic potential?	Yes	No	Yes	No
Can take biopsies	Yes	No	Yes	No
Sedation, analgesia needed?	Mild	Moderate	Heavy	No
Physicians' time	15-45min	Up to 4h (average)	30-75min (endoscopy)	1h
Complications, limitations	Intestinal perforation; sedation; apnea; bleeding; infection	Epistaxis; perforation; bleeding; infection	General anaesthesia; exploratory laparotomy; serosa tears; perforation; bleeding infection	Capsule retention in case of obstruction, stricture, capsule failure

Source: Ali A. Santisi JM, Vargo J. Video capsule endoscopy; a voyage beyond the end of the scope. Cle Clin J Med. 2004; 71; 415-425. Copyright 2004. The Cleveland Clinic Foundation. All rights reserved.

In addition, Table 1-1 shows exactly the advantages of WCE over the rest technologies:

a) Non-Invasive

b) Outpatient

c) Little or no patient discomfort

d) Few or no complications

These four (4) advantages make the WCE procedure patient-oriented and thus the least discomfortable among the endoscopy technologies, while maintaining a high degree of diagnostic successes.

In the future nano-mechanics and nano-technology science will help build robust capsules that will be able to perform therapeutic procedures and biopsies. The Center for Research in Micro Engineering – CRIM Lab (Scuola Superiore Sant´Anna, Pisa, Italy) has published an interesting article on the future of capsule endoscopy [2]. In the appendix of this dissertation I propose a detailed design of a diagnostic capsule and a robotic capsule. Both of them offer better features compared to current capsule technology.

1.3 The need for Computer Aided Diagnosis system

The patient swallows the capsule and as it traverses through the digestive tract it captures approximately 3 frames per second. When the procedure finishes, over 50,000 and sometimes 80,000 screenshots are wirelessly transmitted to a receiver that is worn by the patient. The video is uploaded to a workstation so that physicians may examine it for any intestinal diseases.

In the case of PillCam capsule, Given Imaging provides gastroenterolo-

gists with RAPID Reader software to examine the uploaded video. Although the video is usually 8 hours long, an experienced gastroenterologist can view the video in ~ 2 hours depending on the case. RAPID Reader offers several features that help the medical personnel with the video examination. These features are synopsized below:

- Ability to capture still images
- Ability for multi-image viewing (2-4 sequential frames)
- Adjustable and automatic speed settings
- Zoom
- Quick video preview
- RAPID Atlas, which allows comparison of on-screen case images with a library of reference images
- Rough estimate of capsule location – map

Fig. 1-3. RAPID Reader software (this picture was taken from Given Imaging Inc.)

Although all the above features are very useful, they are mainly tools for video navigation. Additionally, RAPID Reader offers a detection tool – blood detector. The blood detector tool is an artificial intelligence feature that

goes through the video frames searching for red colour. Thus, in case a frame exceeds a certain value of the red component, it assumes there is bleeding and saves the frame for future reference. Although the success rate of the bleeding detector is conflicting and gastroenterologists are not convinced for the robustness of the software-tool, credit should be given since it is a computer-aided diagnosis tool.

Therefore, it can be seen from the above that the need for the development of computer-aided diagnostic methodologies in form of synergies to solve these open research challenges summarized as follows:

a) reduce viewing time

b) detect intestinal abnormalities

In particular, for the challenge a) the synergistic methodology has to search the whole video, and, based on various criteria, will remove redundant frames, thus decreasing the number of frames to be viewed by the gastroenterologist. For the challenge b) the synergistic methodology has to search all the video frames for known diseases and saves only suspicious frames for further evaluation. Therefore for the challenge b) a near automatic detection of abnormalities can be achieved.

Before we move on with the presentation of the various research methods proposed by researchers and our synergistic methodologies on capsule endoscopy, we have to address the weak points of the current WCE that make the development of diagnostic computer based methods quite difficult and challenging:

1) Video resolution is low – 256x256 pixels per frame
2) Capsule may be bouncing and flipping inside the digestive tract
3) Frame rate is low

In particular, the issue 1) makes image interpretation more difficult, since a lot of visual information (details) is lost. The issue 2) creates a problem of a wrong location sequence because, in case of bouncing or flipping, the next frame is not taken necessarily at a forward position but rather at an undetermined place. Finally the issue 3) may leave out important pieces of visual information, which can help with the smart interpretation of the video.

Thus, the overall organization of this dissertation covers a comparative survey on WCE methodologies and the synergistic methodologies used in this research to offer satisfactory and robust solutions to the WCE challenges mentioned above. In particular, chapter 2 presents a brief comparative survey on endoscopic methods. Chapter 3 provides the first-level methods used in a synergistic way to produce complex methodologies needed for the WCE problems. Chapter 4 – 7 describe the synergies of methodologies that produce graphical information for the GUI (Graphic User Interface) engine and finally this information is given to the physician. Chapter 8 – 10: offer synergies of methodologies that help the physicians through case studies (data mining) or through an additional view (3D reconstruction of the digestive tract).

CHAPTER 2: A SURVEY ON WIRELESS CAPSULE ENDOSCOPY AND ENDOSCOPIC IMAGING

Computer assisted technologies, especially image analysis and computer vision methods, can be used to automatically examine WCE videos and extract important features (here pathological) useful for an interpretation of the context of the video frames and significantly reduce the viewing time spent by physicians to examine these videos.

So far, research studies have been done on two key problems of endoscopic image analysis: *malicious tissue detection* and *tissue discrimination*.

The first problem is about the detection of malignant intestinal features like polyps, bleeding, and abnormal regions (tumours). The latter problem is about finding the boundaries of the digestive organs. This enables the medical doctors to easily find areas of interest.

Although some significant research efforts have been proposed for the above problems, endoscopic analysis is still an important field with many open problems, such as recovering of the missing visual information, accurate detection of pathological features, removal of shadows and bubbles from the intestine surface.

It is important to be mentioned here that this chapter offers a survey on WCE and endoscopic imaging by evaluating some representative methodologies. It is also important to be mentioned that WCE technology is used in colonoscopy as well (Given-Imaging PillCam™ Colon Capsule). Thus, re-

searchers have to deal with problems that are common both in small intestine and in colon. That's the reason some other endoscopy methodologies like colonoscopy methodologies are described in this survey. Also this chapter provides an evaluation of each methodology individually for achieving its highest maturity. It does not offer a comparison among methodologies because each methodology has its own approaches and techniques to solving various endoscopy problems and very few methodologies deal with a specific problem. In addition there is no common database with WCE videos for a fair comparison. Thus, this survey is trying to offer credits to methodologies based on that maturity and on those utilized techniques/approaches.

This chapter presents published research that has been done so far on the previous problems. It briefly discusses each methodology, its advantages and disadvantages, and it provides tables indicating the current status of maturity for each method. The main features selected for evaluating each method are presented in tables and these features are: 1) simplicity, 2) reliability, 3) robustness, 4) real problem, 5) performance on real sets, 6) methodology complexity, 7) originality, 8) real implementation, and 9) improvements and others. In other words, the tables offer a general overview of the methodologies in regards to their current stage from the expected realistic completion. For each feature, a weight is given so that the important characteristics of each method are highlighted and placed on the right performance level.

The reason for this survey is not only to be a reference to the image endoscopy, but also to rank how far each methodology is from its final maturity

level based on its potential development.

2.1 Methods and Systems

In this section we provide a classification of the methodologies under evaluation before their full presentation. The classification is based on the theory or experimental techniques used to compose these methodologies for accomplishing their results. Thus, after a thorough evaluation of all methodologies, three (3) major classification categories were created: i) signal-processing methods, ii) color and image processing methods and iii) artificial intelligence methods.

The table below assigns each methodology to these categories:

Table 2-1. Methodologies Classification

	A	B	C	D	E	F/G	H	I	J	K	L	M	N	O
SIGNAL PROCESSING	X			X				X		X	X			
COLOR & IMAGE PROCESSING		X	X		X	X						X	X	X
AI TOOLS		X			X	X	X	X	X	X	X	X		X

Note that capital letters represent the methodologies presented below.

In the Signal Processing category fall all methodologies that involve processing information in the frequency domain. These methods can be Fourier transforms, Wavelet transforms, or band-bass filters.

The Color & Image Processing category includes methods concerning color processing like segmentation, color transformations etc, although these techniques can be viewed as computer vision as well.

The Artificial intelligence category includes methods based on machine learning capabilities, such as clustering and artificial neural networks.

From Table 2-1 it is concluded that nearly all proposed methodologies use artificial tools in order to adapt to the problem they deal with. This shows the level of difficulty and the variety of WCE problems.

Finally, we provide a brief description of each methodology we studied, along with the concluding remarks that helped us to develop the evaluation tables presented at the end of the chapter.

2.1.2 Automatic Detection of Intestinal Juices in Wireless Capsule Video Endoscopy

This methodology [3] deals with the good visibility of the intestinal lumen and wall, which are usually hindered by the presence of intestinal juices mixed up with the remains of food. These intestinal secretions are visualized as a semi-opaque turbid liquid accompanied by the presence of bubbles and other artifacts related to the flux of the different fluids into the gut. As a result, this turbid liquid is interposed between the camera and the gut, obstructing its visualization.

Thus, the main idea behind this methodology dealing with the problem above can be simplified as follows: Firstly, Gabor filters are applied for the characterization of the bubble-like shape of intestinal juices in fasting patients. The use of a bank of Gabor filters has proven to be a very accurate characterization of turbid liquid. Also, a Gabor function can be viewed as a Gaussian function

shifted in frequency to the position (*uo, vo*) and with orientation $\tan^{-1} = \frac{uo}{vo}$. Thus, the Gabor filter has standard deviation of the Gaussian envelope σx and σy, as also *uo* and *vo* parameters. The methodology uses a filter bank of sixteen (16) units with orientations $0^o, 45^o, 90^o, 135^o$ and standard deviation σ = 1, 2, 4, 8. The choice of these values was made through an empirical search. If the size of the filter bank increases, then the detection of false positive areas (intestinal juices) increases respectively. This is undesirable because these false alarms accumulate as noise in the resulting image. Thus, at the start, all sixteen (16) Gabor filters are applied on the video frame and the resulting filtered images are added up according to the following function: $I_n = \sum_{i=1}^{4}\sum_{j=1}^{4} abs\{I_n^{i,j}\}$

The medical specialists selected a set of 100 video frames, in which the area containing the bubbles was delimited by hand and a random set of 300 frames in which no bubbles existed was also selected. In the next step, an exhaustive search was applied over a threshold for I_n (the response images – after the Gabor implementation) to obtain a binary image by defining two exclusively complementary areas in the frame: 1) the region with intestinal juices and 2) the region with no evidence of intestinal juices. The final threshold was searched in an iteratively convergent algorithm which compared the areas obtained by the given threshold value with those described by the specialists. Finally, frames with detected region of bubbles with greater than 50% of the useful visualization area were excluded. The methodology was tested with ten (10)

endoscopic videos. From those videos, 200 frames were chosen and characterized as valid, and 200 frames were characterized as invalid. The method has reached a very high percentage of invalid frames detection – 98% and 95% of valid frames detection. Thus, by using this method, the examination time was reduced by 23% because the medical personnel could skip video frames with no information due to the existence of intestinal juices.

Discussion: *This methodology is quite promising because of its originality on the kind of problem it tries to solve. It can considerably reduce the visualization time spent by the clinical personnel. Their testing pool is not large, but it can be taken into serious account. Although intestinal juices are a portion of the problems in endoscopic image analysis, this methodology may also be used in other bubble-shape textures like food inside the intestine. The main disadvantage of this method is the use of a Gabor filter that their tuning involves empirical training for the right parameters to be found.*

2.1.3 Neural Networks-Based Approach for the Classification of Wireless-Capsule Endoscopic Images

The main idea behind the methodology [4] presented here has also been elaborated in [5],[6].

The purpose of this methodology is to implement a fusion-based methodology to make use of the advantages of color and texture features at the same time. As a first step, texture information is extracted from endoscopic images using nine statistical measures (standard deviation, variance, skew, kurtosis, entropy, energy, inverse difference moment, contrast and covariance) for

each vector of color space RGB and HSV. Thus, the result is 54 features. In this case, the histogram of the original image carries no information regarding relative position of the pixels in the texture. This can fail to distinguish between textures with similar distributions of grey levels. Therefore, there are methods that recognize characteristic relative positions of pixels of given intensity levels. Thus an additional scheme, NTU transformation (texture unit number transformation), is proposed to extract new textures from the texture spectra in the chromatic and achromatic domains, for a selected region of interest from each color component histogram of the endoscopic images. The texture spectra of various image components {*I(intensity), R(Red), G(Green), B(Blue), H(Hue), S(Saturation)*} are obtained from their texture unit numbers.

At the second step, the statistical features (from step 1) are then estimated on the histograms of the NTU transformations of the chromatic and achromatic planes of the image (R, G, B, H, S, V). The six subsystems (R, G, B, H, S and V) that have been developed (step1 & step2), go through a fusion scheme, in which the Extended Normalized Radial Basis Function Network (ENRBF) is used. Although a usual scheme chooses one best subsystem from amongst the set of candidate subsystems based on a winner-takes-all strategy, the current proposed (ENRBF) approach runs all multiple subsystems with an appropriate collective decision strategy. The ENRBF network is incorporated into a multiple classifier scheme, where the structure of each individual (for R, G, B, H, S, V planes) classifier is composed of 9 input nodes (nine statistical features) and 2 output nodes. In the final stage, the nine statistical measures for each indi-

vidual image component-color space, are then calculated through the related texture spectra after applying NTU transformation. The results were very good, giving a final 95.71% accuracy.

Discussion: *The main disadvantage of this methodology is that it does not clearly state what kinds of abnormalities are detected. In addition, their training pool is seventy (70) frames and the testing pool is seventy (70) frames. This small testing pool doesn't strongly support the successful results. The originality of their methodology can be found on the combination of color spaces and textural statistical measures. The main idea behind this method is: If a classifier is "an expert" in some local area of the feature space, then it is wiser to use all classifiers altogether for better results.*

2.1.4 Model of Deformable Rings for Aiding the Wireless Capsule Endoscopy Video Interpretation and Reporting

This methodology [7] deals with the extraction of an estimate of the gastrointestinal tract as the capsule traverses it, so that a map can be created. This map helps physicians to easily go through the video feedback and find areas of interest. The result is a great reduction of the viewing time. In addition, the physician has an estimate of the whereabouts of the capsule. This methodology creates a 2D map of the internal surface of the digestive system and computes a rough estimate of capsule velocity. The velocity of the capsule provides sufficient data for localization of video sequences that show narrow sections of the GI tract, where the capsule stopped or considerably slowed down. The method here uses a Model of Deformable Rings (MDR). MDR consists of nodes connected to form a mesh. The mesh forms concentric rings surrounding the center

of the frame. MDR nodes store information on local image properties, such as RGB color components. The main steps of the method are:

Step 1. The nodes search within their vicinity of the current frame for locations having similar properties using Euclidean distance d. Each node is pushed to such a location.

Step 2. However, the arrangement of the nodes within the MDR mesh has to be retained. This is possible through the elastic matching method that gives a balance between these two effects (Step 2 and Step 1).

Step 3. The elastic matching, which is essentially the combination of the first two steps, is repeated until a state of equilibrium (defined by a threshold) is reached.

The process of elastic matching is repeated for every frame. It must be noted that as the capsule moves forwards, the MDR expands while matching consecutive video frames. If the capsule moves backwards, the model shrinks.

Discussion*: This method presents no statistical results. Overlooking this, according to the authors, it was found that certain characteristics that indicate areas of bleeding, ulceration, and obscuring froth could be recognized within maps. There is no support for this argument. This model deals with images, which are obtained by a smooth forward/backward movement of a capsule. When the capsule moves sideways or jumps, the produced map may be ambiguous.*

2.1.5 *Automatic Classification of Digestive Organs in Wireless Capsule Endoscopy Videos*

As mentioned [8] in the presentation of the previous methodology, one of the main problems in wireless capsule imaging is the creation of a map of the digestive tract. A similar method is the discrimination of the digestive organs such as the esophagus, stomach, duodenum, small intestine and colon. In this case, the physician also has a rough estimate of the digestive tract of the patient. When contractions occur periodically in the digestive tube, they distort the color values. When a capsule enters the next digestive organ, the corresponding color signal has a short-term change, which is the suddenness of the signal change and the increase in energy. Thus, this methodology uses Energy Functions in the frequency domain to characterize contractions extracted from the WCE video. In particular, a video frame is converted from RGB into HSI colorspace. The reason for this choice is because HSI uses components (Hue, Saturation, and Intensity) that are close to the human eye interpretation. Thus, the Energy function of the i-th frame is defined as:

$Ei = \sum_{k=2}^{\frac{N}{2}+1} (| Xi(k) |^2)$, where Xi(k) is the k-th bin of the Fast Fourier Transform of the video frame and $\frac{N}{2}+1$ indicates the frequency $Fs / 2$ where Fs is the sample rate. The High Frequency Content (HFC) is a weighted energy function like the previous one:

$HFCi = \sum_{k=2}^{\frac{N}{2}+1} (| Xi(k) |^2 \cdot k)$ where k is a weight to increase higher frequen-

cies. Finally, the event detection function is defined as:

$$\frac{HFC_i}{HFC_{i-1}} \cdot \frac{HFC_i}{E_i} > T_{event}$$ where T_{event} is an empirical threshold.

However, there is chance that these events might be divided into smaller events because of the local maxima of the color and the threshold. Thus, what it is needed to be done is to merge events into bigger ones if these events follow this simple correlation rule:

$$Corr(Event_i, Event_j) =$$
$$\begin{cases} true\ if\ \sum_{\kappa=2}^{\frac{N}{2}+1}(|Xi(k) - Xj(k)|^2) > Tcorr \\ false\quad otherwise \end{cases}$$

Where, Tcorr is the correlation threshold empirically found.

Discussion*: The results of the method came from ten (10) videos. A high percentage (76%) of detecting correct events-boundaries and a very small percentage of precision (51%) have been achieved. The methodology manages to detect the most of stomach and duodenum, but the accuracy for detecting ileum and cecum were less. Intestinal contractions are quite many and unlike, and both thresholds were extracted empirically. However, this methodology uses the notion of event that consists of a very original attribute and it takes advantage of the high frequency properties of color information.*

2.1.6 Blood Detection in Wireless Capsule Endoscopy Using Expectation Maximization Clustering

The Given M2A newest software version has provided a feature of detecting the luminal blood called Suspected Blood Indicator (SBI), but its accura-

cy is not high enough to replace human examination. The reported sensitivity and the specificity of SBI were 72% and 85%, respectively. In response to this problem, a technique of detecting the bleeding regions automatically utilizing the Expectation Maximization (EM) clustering algorithm and the Bayesian Information Criterion (BIC) has been proposed [9]. The EM technique has become a popular tool in statistical estimation problems involving incomplete data or in problems that can be posed in a similar form, such as mixture estimation [10]. The EM algorithm is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. The ML estimation estimates the model parameter(s) for which the observed data are the most likely to occur. Each iteration, of the EM algorithm, consists of two processes: i) the E-step, and ii) the M-step. In the expectation, or E-step, the missing data are estimated given the observed data and current estimate of the model parameters. This is achieved by using the conditional expectation, explaining the choice of terminology. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. The estimation of the missing data, from the E-step, is used in lieu of the actual missing data. Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration [11].

For this method the key point is to know that there is a certain color range distinguishing blood pixels from non blood pixels. However, most dark pixels in blood overlap with dark pixels in non-blood because dark pixels have no chrominance. Some bleeding types are red, whereas others are blue. Consequently, bleeding pixels with similar characteristics need to be grouped together to extract the unique visual characteristics. The Expectation Maximization

(EM) clustering algorithm is used to train probability models for the blood pixels and non-blood pixels.

Bleeding pixels and non-bleeding pixels were selected to train the system to recognize bleeding. The system is EM clustering that categorizes bleeding pixels and non-bleeding pixels into different subcategories, because as mentioned before the intensity of blood varies (red-dark, red-blue). Each subcategory is modeled by a set of parameters of Gaussian mixture density. Thus, the blood detection is summarized as follows:

Step 1. Remove dark pixels;

Step 2. Conditional probability restriction for selecting the blood candidate pixels;

Step 3. Region size filtering to remove small areas of wrongly identified blood.

Discussion: *In this research work, it is not clarified what kinds of pixels are identified –bleeding pixels or blood pixels. There seems to be a misapprehension. The testing pool is quite small – ten (10) video frames and this may not suitable to support the well achieved results - 92% sensitivity and 98% specificity. However the theoretical background that is used in this methodology is quite promising because EM algorithm is very useful in motion estimation frameworks* [12].

2.1.7 Using Color Distributions to Discriminate Tissues in Wireless Capsule Endoscopy Images

As mentioned in the introduction of this chapter, the detection of vari-

ous organs inside the digestive tract is an important tool for the physician. Of the alternatives that manufacturers have provided, besides SBI, another software feature is called Localization Function, which shows a rough estimate of the route and the relative position of the capsule. In order for the software to work properly the physician must detect and manually mark two locations, the Pylorus and the IV. Thus, a methodology [13] for an automatic discrimination of tissues in the gastrointestinal tract based on WCE images is presented. The methodology can be used for mouth/esophagus and stomach. Note that the same team has previous published two other methodologies for discriminating stomach/intestine and stomach/intestine/colon.

For this method, images are initially converted into HSI color space. Then, the intensity (I) component is removed to achieve intensity invariance and data size reduction. A color distribution histogram is used here, because it is the feature that the physicians need for their investigation. The key point is that the HS histogram of stomach images is shifted towards red compared to the intestine histogram. However, the colon tissue color distribution is very similar to that of the small intestine. Next, data is compressed by applying Discrete Cosine Transform (DCT) followed by Principal Component Analysis (PCA) for better results. Finally, the system is trained to detect mouth /esophagus & stomach, stomach/intestine and intestine/colon using k-Nearest Neighbor (kNN) and Support Vector Classifier (SVC). The two classifiers are compared. The results were very good for the mouth/esophagus and stomach, but not good enough for the intestine/colon. The kNN and SVC classifiers

proved to give very similar results. The SVC classifier managed to provide stomach/intestine median frame error around 50-70 frames.

Discussion*: This methodology uses a very straightforward theoretical background. The use of color information proved to be a very efficient. The main obstacle of the analysis was the existence of artifacts in the tract. Thus, a way to remove them is proposed by averaging a number of neighboring images. Another positive point of this method is that even a high compression ratio doesn't affect the accuracy of the classifiers. The combination of DCT and PCA algorithms works well.*

2.1.8 Color and Texture-based Gastrointestinal Tissue Discrimination

This work [14] is similar to the previous one with the only differences being a) stomach and intestine have been discriminated and b) 28 sub-regions of a frame are used.

2.1.9 Topographic Segmentation and Transit Time Estimation for Endoscopic Capsule Exams

Here, a different approach to the digestive organs detection is proposed. This methodology [15] detects the boundaries of the digestion organs, entrance – stomach (eso-gastric junction), stomach – small intestine (pylorus), and small intestine – large intestine (ileo-cecal valve).

The main tools used here are SVM classifiers [16] and MPEG-7 descriptors. SVMs have an error surface that is free from local minima and has only a global optimum. The features of the images that seed the classifiers are ex-

tracted from the MPEG-7 descriptors, which are 1) scalable color and 2) homogenous texture. The advantage of these two descriptors is that they are widely tested and that they offer color and a texture representation of the image. The main steps of the methodology can be synopsized as follows:

Step 1. Use MPEG-7 descriptors Scalable Color and Homogenous Texture to extract image features.

Step 2. Seed the above features as inputs to Support Vector Machines (SVM) for training. SVM uses 4 different types of kernel functions: a) linear, b) polynomial c) radial-base and d) sigmoid.

Step 3. Assign each frame, in the zone it belongs to, by selecting the SVM classifier with the largest positive distance to its corresponding hyperplane. For this method, the training set was 30,000 frames and the testing set was 60,000 images. Eight-six percent accuracy was reached using a Gaussian kernel as an SVM classifier. Scalable Color outperforms Homogenous Texture. This is quite interesting, because Scalable Color relies on color information. The accuracy percentage cannot be overcome because global model fitting (SVM method) has an "inbuilt" low-pass filter that reduces accuracy. Another problem is that the automatic system, given the fact it is based on color and texture features, annotates as a valve the moment the capsule crosses an area with completely different characteristics, but when the capsule bounces, then it annotates these bounces as multiple valves.

Discussion*: The major advantage of this method is the utilization of SVMs for training the system. These classifiers prove to be better than the Euclidean and Mahalanobis*

distances, because they are free of local minima. In addition, the MPEG-7 descriptors that are used here are also efficient due to the fact that they are well established and represent information about the content, not the content itself. The method carries information about low level features in the content (colors, textures etc.)[17].It must be noted that the use of a radial-base function (a Gaussian function) as SVM classifier gave the best results. The combination of Scalable Color and Homogenous Texture descriptors also proved to work better.

2.1.10 Neuronal Networks and Textural Descriptors for Automated Tissue Classification in Endoscopy

Detecting malignant regions in endoscopy video sequences encounters several problems due to the time varying nature of the process and variable perceptual conditions (shading, shadows, lighting conditions, reflections etc). The methodology [18] here deals with issues by extracting texture features from video frames using a) co-occurrence matrices and b) 2D discrete wavelet (DWT). From the co-occurrence matrices were extracted four (4) texture features: a) energy-angular second moment, b) correlation, c) inverse difference moment and d) entropy at four (4) different angles: a) 0^0 b) 45^0 c) 90^0 d) 135^0. Thus this resulted into sixteen (16) features. Applying DWT to each frame provide four (4) wavelet channels: one approximate and three detailed. The final features are 3x16 = 48, that consist the input to the neural network. The method here uses various kinds of neural networks for training, testing, and comparison : (1) offline adaptive back propagation (BP) algorithm with 100 multilayer

perceptrons (MLPs), (2) offline adaptive BP with 1000 MLPs , (3) offline Levenberg-Marquardt (LM) with 1000 MLPs, (4) offline Rprop algorithm, scaled conjugate gradient (SCG) algorithm and LM algorithm with 1000 MLPs and 11 hidden neurons, (5) online Rprop algorithm, (6) online BP, (7) online ABP, (8) online evolution, and (9) ensemble-based. The last algorithm - ensemble-based - is the main proposed training method. It is a combination of several individual networks combining a variant of the stochastic gradient descent with a differential evolution strategy to readapt the neural network to modified environmental conditions [18].

Discussion: *This methodology is a straightforward approach to the problem of malicious tissues. First texture features are extracted from the endoscopy images. Then these features are seeded to neural network training. Neural networks offer the advantage of generalization, working under incomplete information and noise. All of the training methods gave very good results (above 80%). Much credit must be granted for the online ensemble-based training method. It gave 93% classification success. However, throughout this research work it is not clarified what kind of malignant tissues this methodology detects. Online training is very promising technique, because it has the feature to adjust neural network in changing conditions, but it has the disadvantage of having to be retrained over and over in order to be more suitable. In addition online training requires medical personnel to be trained for this purpose.*

2.1.11 *Neural Network-based Colonoscopic Diagnosis using Online Learning and Differential Evolution*

This methodology [19] here is also based on various methods of offline and online training Neural Networks [18].It gives more emphasis on online training with online back-propagation (BP) learning and online differential evolution strategy (DE). As it was mentioned previously, Neural Networks are good techniques because they offer very strong features for real problem solving. They learn from experience, and have generalization and the ability to cope with uncertainty in distorted or noisy images to some extent. Online training, compared to offline training, offers the advantage of adjusting the Neural Network over and over by taking into account previously acquired knowledge (memory). The authors claim that online training is more efficient because it helps the process to escape from local minima and provides an approach to learning in non-stationary environments. This positive view of online training can be considered as a disadvantage, because here an ideal tradeoff must be found between the newly acquired knowledge and its memory. Thus, adaptation must be as smooth as possible. In this case, texture features are extracted from colonoscopy images using co-occurrence matrices (energy-angular second moment, correlation, inverse difference moment and entropy), and online BP learning is used to initialize the population of the Differential Evolution (DE), as BP provides an initial approximation of the solution. Online BP is widely used in online Neural Network training, however here it suffers from excessive sensitivity to learning parameters. In addition, in [20] it was experimentally

reported that DE strategies have shown good convergence properties and outperform other evolutionary algorithms and annealing methods. DE can handle 1) non-differentiable, 2) non-linear and 3) multimodal objective functions efficiently and require few easily chosen control parameters [21]. Basically, DE adds the weighted difference between two population vectors to a third vector. This way, no separate probability distribution has to be used, which makes the scheme completely self-organizing [21].

The major steps of the methodology here are: 1. An n-D weight vector is used as initial population offered by BP; 2. The weight vectors are generated by the combination of vectors randomly chosen from the current population (mutation); 3. The resulting vectors are mixed with a predetermined target vector producing the trial vector; 4. This trial vector is accepted for the next generation if and only if it reduces the value of the objective function. This last operator is referred to as a selection.

Thus, the DE presentation offline training is presented and all the results from the various methods were compared. The best results came from the Rprop offline training algorithm, because offline training uses the true gradient of the error function as it exploits information from all the training patterns. The online BP with DE provided the second best results, because it works with an instantaneous approximation of the true gradient and because it uses information only from one pattern per iteration. On the other hand, the online method has the advantage of re-adapting to modified conditions, while R*prop* requires all information a priori.

Discussion*: This methodology is based on the Differential Evolution technique (1990) and offers promising results. Perhaps it does not produce the same successful results as offline Rprop for the reasons that are mentioned above; however, its characteristics and advantages have to be seriously taken into account for future research in the area of endoscopy image analysis.*

As it was mentioned in the introduction, the purpose of this chapter is to offer a survey on endoscopy methodologies besides WCE methodologies due to the fact on similarities on problems and objectives in both areas. Thus, the following research methods presented here are used for identifying colorectal polyps in colonoscopy video frame sequences. Polyps are characterized from: 1) their color, 2) appearance of their mucosal surface, 3) presence of ulcers, 4) their bleeding tendency and above all, 5) the presence of pedunculus (pedunculated or nonpedunculated).

2.1.12 Computer-Aided Tumor Detection in Endoscopic Video Using Color Wavelet Features

The main concept of the methodology [22] discussed here can be briefly described as follows: A novel feature extraction technique was used based on a discrete wavelet decomposition that was applied on different color spaces. Then statistical analysis of the wavelet coefficients associated with the color bands was used. The wavelet features, named Color Wavelet Covariance (CWC), were based on second-order textural information estimated on the do-

main of the discrete wavelet decomposition of each color band of a video frame. The methodology is as follows:

Step 1. RGB video frames are converted into Karhunen – Loeve color space resulting in three color channels. This diagonalizes the covariance (or optionally, the correlation) matrix of the colors. It performs very well for color texture characterization [23].

Step 2. Since the size of the lesions varies, a multi-resolution approach plays an important role. Due to the fact that texture is essentially a multi-scale phenomenon, the Discrete Wavelet Transform (DWT) works well here because it offers spatial/frequency representation. As it is known, when the decomposition level decreases in the spatial domain, it increases in the frequency domain giving zooming capabilities. So the DWT transformation consists of: 1) convolution of the image with a lowpass filter H, 2) convolution of the image with a bandpass filter G, 3) repeatition of the previous convolutions to decompose image at several scales. The result is a set of the low resolution image Ln along with all the detailed images $D_1 \ldots D_{3n}$, where n is the scale.

Step 3. Co-occurrence matrices were used on images from the previous step to extract second order statistical measures, such as: 1) Second moment; 2) Correlation; 3) Inverse difference moment; 4) Entropy. Two more features were used, i) Contrast and ii) Dissimilarity, but their results weren't satisfactory, so they were removed from the process.

Step 4. The covariance among these different statistical values was applied between the various color channels of the examined region in order to

rank changes. In other words, using the covariance between two features, a measure of their "tendency" to vary together was given.

Step 5. Finally, linear discrimination analysis (LDA) was used for classification and evaluation of the proposed features.

In this methodology, a very good percentage of sensitivity (93.6± 0.8%) and specificity (99.3± 0.3%) were achieved. It needs to be mentioned here that a 128x128 pixels window was used, because the smaller the window size set the bigger the mean classification error (MCE) and the uncertainty. This means that a large population of pixels is required to characterize tumor regions using second-order statistical features in the wavelet domain.

Discussion*: The performance of this methodology was very efficient reaching high percentages of success and maturity. It combines color textural information extraction through DWT of endoscopy images. The K-L color space has proved to behave better in colonoscopy images. However, it is not argued quite well why DWT should be used since low-frequency images don't carry major texture information. It was also indicated that the use of the proposed classification method – LDA - is not quite robust and alternative classification should be found.*

2.1.13 CoLD: A Versatile Detection System for Colorectal Lesions in Endoscopy Video-frames

This work is an incremental improvement of a methodology discussed in the previous paragraph 2.1.12 using a graphical software implementation. This software [24] can be used and trained by the medical personnel to detect precancerous polyps in colonoscopy images. The new part here is only the classifi-

cation scheme, an Artificial Neural Network (ANN) that can handle nonlinear multidimensional input features and they have the advantage of generalization. The NN used in the software system consisted of an online BP with 48-X-2 MLPs, where X is the number of hidden neurons. The relationship between X and time was linear. Thus, the modified methodology is presented below:

1) Take a region of the image (square window).
2) Apply DWT to this region resulting into 3 bands (bands 2, 3 and 4)
3) Calculate co-occurrence matrix for each band resulting 16 statistical measures for each band (3x16 = 48 textural value vector)

This above procedure is applied either for training or testing. In the training case, the training set is calculated on a limited set of regions that are randomly selected from the image. In the testing case, overlapping windows are used. Any examined area is assigned to a class corresponding to the normal or abnormal case, and this information is stored as the 49th component of each 48th vector.

Isolated misclassified vectors or group of vectors may appear as spots or small formations within the output processed image. A post-processing stage is applied to cut out these "errors." The medical personnel train the neural network by tuning the MLPs. The software is user-friendly.

Discussion*: Continuing their good research performance done in [19], the researchers improved it by offering user-friendly software based on their solid methodology and by changing the LDA classification scheme to the better familiar and trustworthy Neural Network. The use of online BP training has the advantage of offering better results than the offline training in general (shown in II-9), however it lacks of versa-*

tility as it needs training and retraining. It must be noticed that the average success rate exceeds 95% without retraining the NN. The methodology also has fast training and evaluation times. The probability of false alarms is higher than the probability of neglecting regions corresponding to suspicious lesions, but this is not a major disadvantage.

2.1.14 Classification of Endoscopic Images Based on Texture and Neural Network

This methodology [25] proposes different ways of detecting malignant regions in endoscopy images, such as ulcer or coli. A Local Binary Pattern (LBP) is used for texture feature extraction and then a log-likelihood-ratio, called G-statistic, is used for the evaluating LBP distributions. The basic idea of LBP is a binary code that describes the local texture pattern. This code is built by thresholding a neighborhood by the value of its center. Through its recent extensions, the LBP operator has been made into a really powerful measure of image texture, showing excellent results in terms of accuracy and computational complexity in many empirical studies. This work uses, however, a modified LBP method combined with the intensity I of the image to generate more stable results for endoscopic images. Note that the original LBP method doesn't use contrast and intensity of the texture.

The G-statistic, which is a modification of Kullback's criterion, shows the probability that two sample regions are from the same population. The higher the value of G-statistic, the less the probability they are. Also, a Self

Organizing Map (SOM) is used for the classification. SOM is an unsupervised process and this is a great advantage given the complexity of endoscopic images.

Discussion*: There are no results given for this methodology. Their absence does not allow an accurate maturity ranking. However the use of G-statistic as evaluation and LBP as texture extraction scheme compose an original concept.*

2.1.15 A Methodology for Detecting Blood-based Abnormalities in Wireless Capsule Endoscopy videos

This methodology [26] deals with bleeding detection in WCE videos in a more sophisticated way. Past research works have focused only on identifying bleeding. The present methodology introduces a novel categorization consisting of three (3) cases of blood-based abnormalities: a) Bleeding, b) Angioectasia and c) Erythema. It utilizes a mathematical transformation as a color space transformation, known as Karhunen – Loeve to make R, G, B values as much uncorrelated as possible. More specifically, as described in [27], let S be the region to be segmented and Σ be the covariance matrix of the distributions of R, G and B in S. Let $\lambda1$, $\lambda2$ and $\lambda3$ be the eigenvalues of Σ and $\lambda1 \geq \lambda2 \geq \lambda3$. Let $Wi = (W_{Ri}W_{Gi}W_{Bi})^t$ for i=1,2,3 be the eigenvectors of Σ corresponding to λi, respectively. The color features X1, X2, X3 are defined as: $Xi = W_{Ri} \cdot R + W_{Gi} \cdot G + W_{Bi} \cdot B$ ($\|Wi\| = 1$, i=1, 2 and 3). After applying this transform onto WCE frames, blood-based cases appear more distinct from the rest image. A complex segmentation process [28] based on fuzziness extracts the regions of a WCE

frame. After this step, a merging procedure is carried out based on Local–Global (LG) graph scheme [29] to combine similar regions into bigger ones. In such a way, small blood-based segments are not discarded. This scheme uses color information, geometry information and segment interrelation of each region to achieve the merging process. Finally, an experimental color threshold is applied to keep only blood-based regions.

Discussion*: The methodology introduces a novel categorization of blood-based abnormalities and for that credit should be given. However, it is not mentioned how these three types of abnormalities can be detected or categorized. The use of the Karhunen – Loeve transformation seems promising as a tool compared to other color spaces in WCE field. The evaluation of the methodology lacks of serious number of results so the degree of robustness is unknown. It should be noted that this methodology has been developed under Windows®.*

2.1.16 A Neural Network-based Detection of Bleeding in Sequences of WCE Images

This methodology [30] copes with the problem of bleeding detection in WCE images. As it was mentioned previously, Given-Imaging© has provided its software package with a bleeding detection feature. However, its detection rates vary from 21% to 41%. The present methodology detects bleeding successfully at 80% with potential for high rate of success. The main concept used in this methodology is color features. It applies a Color Adaptation approach by using a reference color value of the image. Color adaptation is carried out by the Cone Response Transform and the reference color is detected using a

Neural Network. Color histograms are produced first for the whole image and second for the designated blood area. Both histograms are input and output in the training stage of NN. Image color is detected in the testing phase after color adaptation and estimation of deviation from the gray value axis in the RGB cube. Color adaptation can be considered as a dynamic mechanism of the human visual system to optimize the visual response to a particular viewing condition. Chromatic adaptation is the ability of the human visual system to discount the color of the illumination and to approximately preserve the appearance of an object [31].In the bleeding detection problem, this helps to visualize the bleeding areas better. However the reference color used by color technique needs to be extracted by the image. This is done by a NN through the histograms of the whole image and the bleeding areas. The selection of bleeding areas is done manually by the physician so that the NN can be trained.

Discussion*: This methodology is based on color adaptation*. *The key point is how to increase the visualization of bleeding areas. Color adaptation offers this feature. The testing samples of this method and results are very small in size. However the fact that there is a developed software tool that implements this methodology adds credit.*

2.2 Selection of Features for Evaluation

In this section we provide a number of features for the progressive evaluation of all the methodologies mentioned above and ranking their level of maturity. These features were selected by the research group of the ATRC after useful discussions with gastroenterologists, software developers, and manufacturers. Thus, the features selected here are as follows:

Table 2-2. Proposed Features

Simplicity (F1)	The methodology is simple.
Complexity (F2)	It reflects the computational time and "p" for the methodology to produce the desirable output or assigned task.
Originality (F3)	The methodology is based on original algorithms and or mathematical formulas or the synergistic combination of simple methods composing a new original method.
Real Application (F4)	The methodology is applicable to solve real problems of high importance.
Real-Time (Processing Time) (F5)	The methodology can be implemented in real-time. It can produce results as soon as the input data is fed to the methodology.
Product (F6)	The methodology is a commercial product.
Computational Requirements (F7)	The computational power required by the methodology to produce desirable results.
Friendliness (to user) (F8)	The methodology offers a user-friendly interface so that the physician can work easily and sufficiently with it.
Performance and Test on Real Cases (F9)	Sufficient results and performance statistics are presented to back up methodology's efficiency on real cases.
Reliability (F10)	The methodology produces reliable results.
Robustness (F11)	The methodology produces acceptable results under any circumstances.
Training Time (preprocessing stage) (F12)	The methodology requires extra time for preprocessing the data set (i.e. NN training).
Cost (F13)	The amount of the manpower needed to create the product
Invasive (F14)	The methodology requires a preprocessing stage, such as a surgical operation, in order to produce its results.
Categories of Pathological Abnormalities (F15)	The methodology can detect one or more pathological abnormalities.
MD interaction (F16)	The methodology requires physicians' involvement to produce results.
MD and/or FDA approval (F17)	The methodology has been approved by the medical community and/or by the Food and Drugs Administration.
Further Improvements (F18)	The methodology has the potential of further enhancement.
Implementation (F19)	The methodology has already been implemented in a medical environment.
Prototype (F20)	The methodology has been successfully implemented at the experimental stage and produced desirable results.

2.3 Informative Tables

In this section, we provide informative tables that offer information of

the current status and the potential capabilities for each of the methodologies. In particular in Table 2-3, all methodologies are listed and represented by symbols in alphabetic order (from A-to-O, top row) as they were described above. The first column of the table represents the features expressed by their symbols (from F1 to F20). In each cell of the table, a "1" is designated if the particular feature is available in the methodology, or a "0" if the methodology cannot carry this feature, or "X" if not enough information is available regarding this feature, or "NA" if no information is available about it.

Table 2-3. Feature availability in each methodology

	A	***B***	***C***	***D***	***E***	***F***	***G/H***	***I***	***J***	***K***	***L***	***M***	***N***	***O***
F1	1	0	0	1	1	1	0	0	0	0	1	X	0	1
F2	0	1	1	0	0	0	1	1	1	1	0	X	1	0
F3	1	X	1	0	1	0	1	1	1	1	0	1	1	X
F4	X	1	1	1	1	1	1	1	1	1	1	1	1	1
F5	X	X	X	X	X	X	X	X	X	X	X	X	X	X
F6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F7	1	1	1	1	X	1	1	1	1	1	1	1	1	X
F8	NA	NA	NA	NA	NA	NA	NA	X	X	NA	NA	NA	1	X
F9	1	0	0	0	0	1	1	1	0	1	1	0	0	0
F10	1	1	X	X	0	1	1	1	1	1	1	1	1	1
F11	X	1	0	X	0	0	0	0	1	X	1	X	1	X
F12	1	0	0	0	0	1	0	1	1	1	1	1	1	1
F13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F15	0	X	0	0	1	0	0	1	1	1	1	1	1	1
F16	0	0	0	0	0	0	0	1	1	0	1	0	1	1
F17	0	0	0	0	0	0	0	0	0	0	X	0	1	1
F18	1	0	1	0	0	1	1	1	1	1	1	X	X	X
F19	0	0	0	0	0	0	0	X	X	X	1	0	X	X
F20	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**The appearance of NA features serves the purpose of a future maturity evaluation which might include information for those features*

Some useful conclusions can be identified from Table 2-3. Almost all methodologies deal with problems of great importance in WCE. All methodol-

ogies are software based and none are hardware implemented. All of them, except L methodology, are implemented in experimental stage showing that there is more room for further improvements. One important conclusion that may also be derived from this table is that there seems to be a lack of strong collaboration between physicians and engineers. In such a medical field like WCE, the role and presence of a team of physicians plays an important task in coordinating research. In such sense, the only true effort comes from methodology L, which since it is a Windows based software, offers the capability to be tested in real-time and full scale by physicians.

After several meetings with physicians and patients from Digestive Association Inc. in Dayton, OH, we came up with Table 2-3. In Table 2-3, each feature is assigned a weight, which reflects its importance from the engineers' and developers' point of view (*weights table*). In Table 2-4, the weights provide a degree of importance of each feature from the perspective of the physician, patient, and manufacturer. For example, robustness is a much more important feature than real-time processing, because the physician cares more about successful results than real-time processing. The group of engineers at the ATRC who evaluated the methodologies using these features graded each methodology if that was possible, or else left out empty with a value (x_i) between 0 and 10 (*scores table*). If one characteristic is not graded – *left empty* this means that the characteristic is not available (NA) or we did not have enough information (gap). Finally using the formula $T = \sum_{i=1}^{N} w_i x_i / N + b$, a total score of each me-

thodology was calculated, where "i" refers to a specific characteristic, N is the total number of characteristics for each system and b is bias (for now, b=0).

Table 2-4. Weight assignment of features

	PHYSICIAN'S PERSPECTIVE	PATIENT'S PERSPEC-TIVE	MANUFACTURER'S PERSPECTIVE	AVERAGE
F1	1	1	9	3.7
F2	1	1	9	3.7
F3	1	1	7	3
F4	10	10	9	9.7
F5	3	6	8	5.7
F6	10	10	10	10
F7	3	1	8	4
F8	9	7	9	8.3
F9	10	10	10	10
F10	9	10	10	9.7
F11	10	10	10	10
F12	8	1	9	6
F13	4	2	10	5.3
F14	9	10	7	8.7
F15	10	10	10	10
F16	10	2	9	7
F17	9	9	10	9.3
F18	6	3	10	6.3
F19	9	9	9	9
F20	5	5	10	6.7

Manufacturer's perspective combines evaluation by the manufacturer as well as by research engineers.

Table 2-4 shows there is strong agreement from all three (3) parties on common obvious feature aspects, such as reliability, robustness, testing, MD/FDA approval, and non experimental implementation.

There are also features that one party weighted much higher than the rest did. So for example, it is reasonable that from the manufacturer's perspec-

tive, complexity of a system is of high importance because of cost inefficiency, whereas for a physician or a patient it matters little since they care far more for the accuracy and reliability.

Table 2-5. Methodologies grading for each feature

	A	B	C	D	E	F/G	H	I	J	K	L	M	N	O
F1	8	5	2	8	7	4	2	4	5	2	5	5	2	6
F2	2	7	8	3	4	7	8	7	7	8	6	5	9	4
F3	6	7	9	6	6	7	7	7	6	9	5	6	9	7
F4	6	9	9	6	7	6	6	8	9	9	8	4	9	7
F5	X	X	X	X	X	X	X	X	X	X	X	X	X	X
F6	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F7	6	8	8	6	7	8	9	8	8	8	8	8	8	7
F8	NA	NA	NA	NA	NA	NA	NA	3	3		7		8	3
F9	8	7	2	2	2	7	9	8	8	9	9	6	8	2
F10	6	8	7	4	6	7	7	8	8	9	9	6	9	8
F11	6	7	7	6	6	5	6	8	6	8	8	6	8	8
F12	5	1	1	1	1	5	1	6	6	4	7	6	7	5
F13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F15	1	7	1	1	4	1	1	1	7	7	8	6	8	4
F16	1	1	1	1	1	1	1	7	7	1	8	1	1	6
F17	1	1	1	1	1	1	1	1	1	1	6	1	4	1
F18	7	1	6	4	3	5	5	1	5	7	8	2	6	6
F19	1	1	1	1	1	1	1	5	5	3	8	1	1	6
F20	1	1	1	1	1	1	1	2	2	1	7	1	1	5
TOTAL	462	527	437	343	400	453	460	613	688	639	894	455	730	613

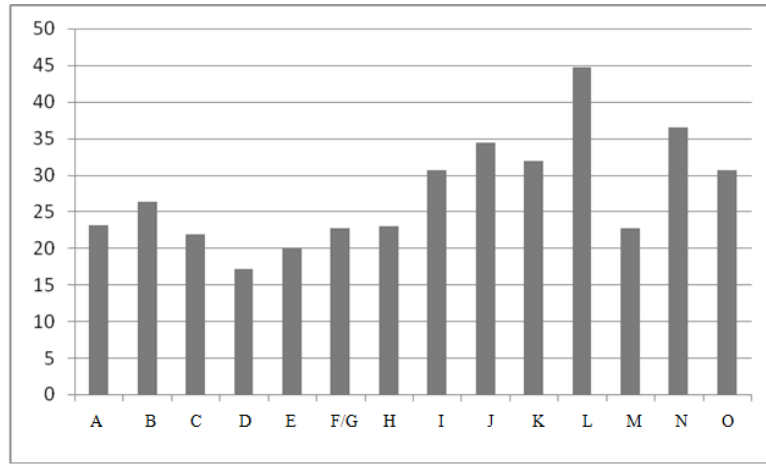


Fig. 2-1. Final grading of methodologies: The ranking of maturity for each method

In Table 2-5, scores were assigned after having been examined thoroughly by the engineering team in the ATRC lab in coordinance with the physicians at Digestive Association Inc., as mentioned before. In addition, the vertical axis represents the level of maturity expressed on a scale from 0 to 50, with 50 maximum. The scoring of methodologies in the case of the complexity feature might seem dubious at first, since each methodology deals with a different problem and no software testing took place – we didn't have the algorithm codes to run them. However, the team put a great effort to assign scores to the complexity feature based on how complicated the methods were as used in each methodology.

As it was mentioned in the introduction of this chapter, the survey aims to offer to the researchers and readers pieces of valuable information regarding the level of maturity of some representative WCE and Endoscopic Imaging methodologies for individually reaching their goals. Thus, Fig. 2-1 shows the levels of maturity for each of the methods used here for evaluation. The methodology represented by letter L has reached a very good level of maturity and performance as mentioned before. The methods I, J, K, N, and O have reached a satisfactory level of maturity, but they have space for further improvements. The rest methods present good features but they have more room for improvements. It must be mentioned that for the above maturity evaluation of the presented methodologies, we did not further categorize the methods due to a small number of them per category and the difficulty to derive specific results from them. Thus, our evaluation is based on a general categorization by using a

criterion of contribution of methodologies to the WCE.

2.4 Conclusions

In this chapter an informative survey on methodologies that deal with endoscopic imaging was presented. The survey was based on a variety of features selected by the authors, a group of MDs, and manufacturers in order to evaluate the efficiency of the methodologies. There were no outstanding methodologies, as shown throughout the survey, which meet all requirements to achieve maximum efficiency. In fact it is very hard for any methodology to satisfy all the aforementioned features because of the great variety in WCE problems. In particular, there are many issues in WCE image analysis that need to be solved such as bleeding, peristalsis, illuminations, abnormal tissues etc., as well as the creation of an accurate map of the digestive tract. It is unfeasible for only a single methodology to deal with all these issues. Perhaps, an integrated methodology containing various components with specialization in certain issues would perform better.

This survey had a goal to present various methodologies that have been published in the field of endoscopy image analysis. The survey didn't have the intention of rewarding the best methodologies by ranking them. The reason is that each methodology offers a different approach to a specific WCE problem, and it is this diversity of viewing that was rewarded. More specifically the survey's intention was to address the present stage of each methodology and indicate its current potential improvements to achieve better performance and

reach the highest level of maturity.

As far as in which direction future research in WCE technology should head to, there cannot be a single definite answer. However, having a closer look on WCE problems, one could indicate that serious advances should take place in hardware to improve the quality of captured video (higher resolution and higher frame rate), to add more sensors (position tracking, pH measuring, temperature measuring), and elongating life battery. Next generation capsules should employ sophisticated tools to perform biopsies and to be fully controlled in real-time, or not by physicians or software based systems.

Present and future efforts in software methodologies, which this survey described, will need to reach a point of better performance regarding with image processing and computer vision techniques, but further improvements may not be feasible as long as there are no advancements in the hardware and MDs essential participation. Future capsule generations will provide software engineers with better quality and larger quantity information, letting them create smarter software systems to deal with current unsolved issues and to offer more capabilities.

CHAPTER 3: METHODOLOGY AND INCORPORATED METHODS

In this chapter we briefly present the overall configuration of the architecture (scheme) that generates the synergies of methods from single ones. In addition, it offers a short description of the singular methods used here in the synergies for better understanding of the synergies themselves.

3.1 The overall architectural scheme of the proposed synergies of methodologies for WCE videos

The general architectural scheme of the synergistic methodologies is given in Fig. 3-1. The scheme of the methodology follows a top-down approach. At the very top of the scheme is the input of the system, which is the raw video data. At the very bottom the output of the methodology is presented to the gastroenterologist to make the final decision. Thus, the information transverses through layers to be processed and it is transformed to a meaningful output that can be understood by the physician. Therefore the decision making part of this scheme is the gastroenterologist/authorized medical personnel themselves.

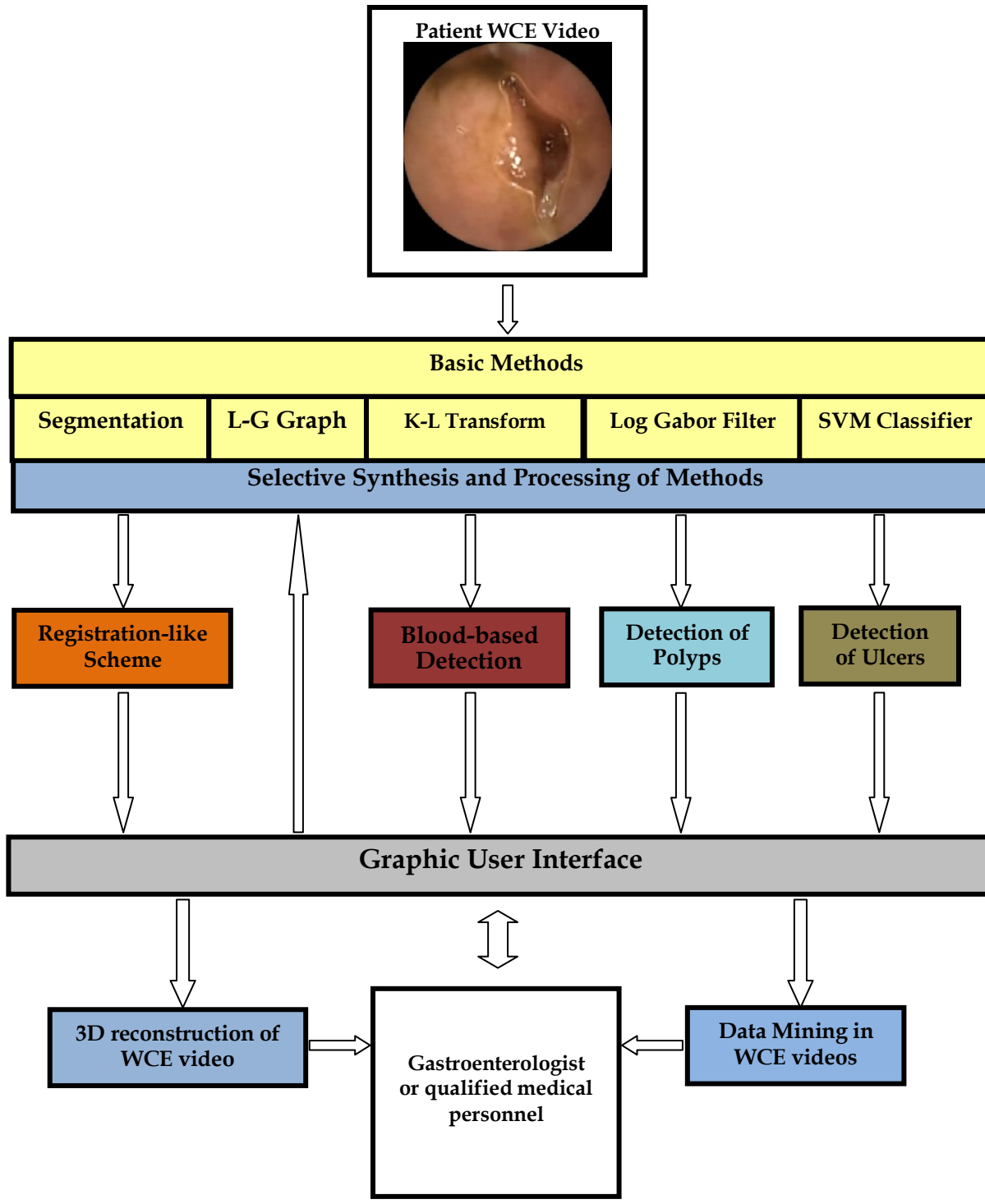


Fig. 3-1. Oveview of the proposed CAD system

3.2 Fuzzy Region Segmentation (FRS)

The FRS segmentation scheme used here is based on a previous work [28]. Here a brief description of this FRS segmentation scheme is given as graphically is presented in 0.

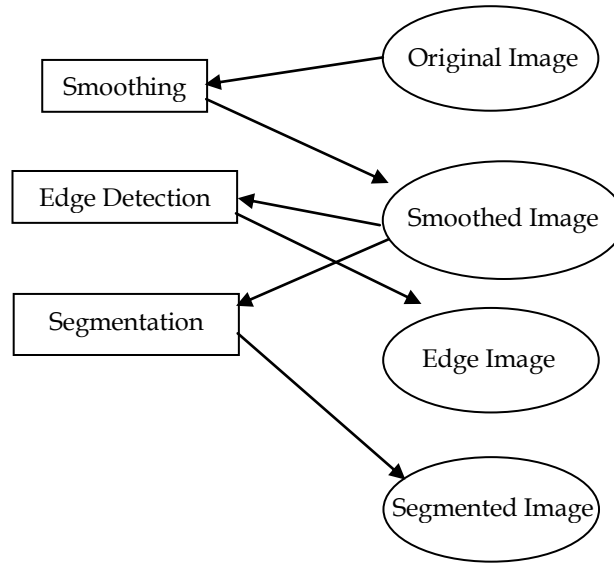


Fig. 3-2. Fuzzy Region Segmentation scheme

Initially a smoothing operation is applied to remove noise from the frame. This operation attempts to preserve the location of edges by only smoothing areas where local contrast (within a small neighborhood) is fairly low. The edge information is utilized from the smoothed input image. The segmentation algorithm uses edge information and the smoothed image to find segments.

The processes involved in this segmentation procedure are as follows. The edge detection step creates segments that are surrounded by edge pixels or the image boundary. A solid segment is defined as a set of pixels completely

surrounded by edge pixels belonging to only one object. To find a solid segment, the image is first scanned for the first non-edge pixel. This pixel is used as a growing seed. In the growing process, a pixel can grow recursively in four directions (left, right, up and down), and merged with the seed if the growing condition is met.

The growing process for a segment ends, if the grown segment is not sufficiently large (i.e. consisting of a threshold quantity of pixels as determined by the image's histogram), and the segment is removed. This removal is due to the assumption that the segment found may actually be a part of a larger segment, but was falsely separated because of noise. All the pixels of the segment are categorized in groups of similar color. The average color of each group is then calculated. The number of pixels in each group gives a weight that is used to calculate the average color of the segment. The segment is painted with this color.

The next step is the expansion of the big solid matte segments found during the previous phase. The expansion procedure is performed within three sub-phases, during which, the segments are enlarged to include matte pixels with the already delineated matte segments. Larger segments are expanded first. During these sub-phases, each segment is expanded (surrounding pixels merged with the segment) only if the resultant segment is homogeneous. Homogeneity degree is defined as a fuzzy term is high if:

1. The absolute similarity (similarity between a pixel's color and the

segment's color) is high

(Eq. 3-1) $|C(P_n) - C(R_i)| > \tau_1$ where $C(P_n)$: color of pixel n, $C(R_i)$: color of segment **i** or

2. The local or relative similarity (similarity between the next and the previous pixel's color in the growing direction) is high

(Eq. 3-2) $|C(P_{i,j}) - C(P_{i,j\pm2})| > \tau_2$ or $|C(P_{i\pm2,j}) - C(P_{i,j})| > \tau_3$ where $C(P_{i,j})$ is the color of pixel i,j

where C(Pn) is the color of current points, and C(Ri) is the color of the region τ1, τ2 and τ3 are selected thresholds. This condition is added to include areas that gradually change in shade while avoiding crossing the object boundary.

Color contrast can be measured by computing the difference between two color vectors and obtaining its magnitude. As shown in (Eq. 3-3), the square of the Euclidean distance is used to calculate the color contrast between two-color vectors v and w in this approach:

(Eq. 3-3) $\mathbf{Contrast\ (v,\ w) = (Rv - Rw)^2 + (Gv - Gw)^2 + (Bv - Bw)^2}$

where R, G and B are the three-colour components (Red, Green and Blue). Similarity membership functions are calculated using the equation shown in Fig. 3-3.

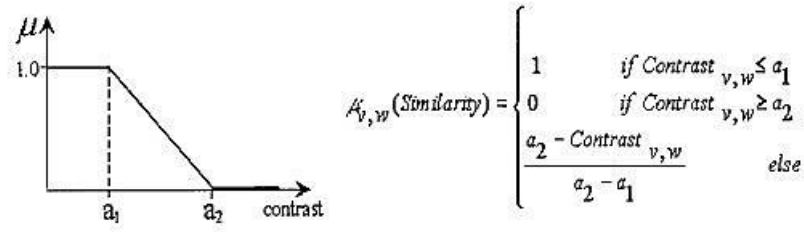


Fig. 3-3. Similarity Membership Function

The Farness measure is used to further expand segments. An unassigned pixel can be close (not far) to a neighboring segment in two senses: 1) close in the spatial domain (physically close); or 2) close in the cluster domain of the color cube (almost of the same color). The degree of farness of a pixel to a neighboring segment is defined as a product of these two measures. Specifically, the degree of farness for any given pixel is the absolute color contrast multiplied by the geometric distance (in pixels) between the given pixel and the segment border. During this expansion, the degrees of farness for a candidate pixel are computed for its three most closely neighboring segments. The closest segment to the pixel is the one having the lowest degree of farness. If this lowest value belongs to the expanding segment, and is lower than a fixed threshold, the pixel is merged with that segment.

3.3 Local Global Graph

One of the methods to describe the shape of an object in an image is the graph methodology [131]. The purpose of the graph methodology is to model the geometry of objects in mathematical forms. Having accomplished this, it is

much easier to extract useful information of the objects. The Local Global (L-G) graph is such a method (Fig. 3-4) used here. The main components of the L-G graph are:

a) The region or local graph that represents the information related with color, texture, shape, size etc for each image region;
b) The skeleton graph that provides information about the internal shape of each segmented region;
c) The global graph that represents the topological relationships among the segmented regions for the entire image;

Thus, the basic idea behind this L-G graph based method is the local and global geometric representation of the image features and their relationships. The selection of the graph structure for representing the information extracted from an image is important for three reasons:

1) It is a generic and very flexible information representation scheme
2) It is very robust and computationally not very expensive
3) It is well known that similarly humans visually observe the environment and extract main features or landmarks and inter-relate them for a better representation in their minds.

From the three components mentioned above, we can see that component 1 includes not only geometric information but also color and texture. Firstly, this has to be noted for future reference and secondly, it emphasizes the generality of this method.

The Region or Local Graph holds information of a contour – line of an

image region after segmentation.

$$G = N_1\ a^{c}_{12}\ N_2\ a^{c}_{23} \ldots\ldots N_k a^{c}_{k1} N_1 \lozenge N_i a^{p}_{ij}\ N_j\ \lozenge \ldots\ldots \lozenge N_n a^{rd}_{nm} N_m$$

where ◊ represents the graph relationship operator, and each Ni maintains the structural features of the corresponding line segment, thus, N_i = { orientation (0), length (le), curvature (cu)}, and a_{ij} holds the relationships among these line segments, thus, a_{ij} = {connectivity (c), parallelism(p), symmetry (s), relative magnitude (rm), relative distance (rd), etc. }

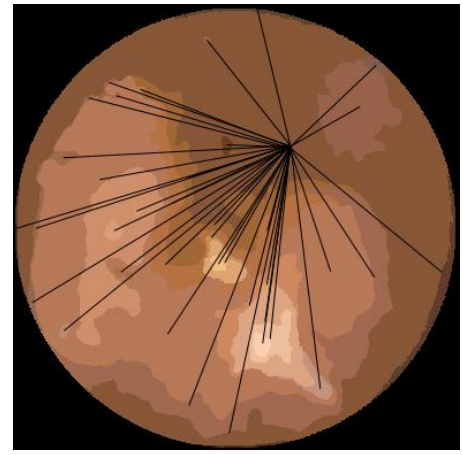


Fig. 3-4. The L-G graph visual representation for a single region (a few global graph connectivities are shown for the sake of the image complexity)

The Skeleton graph is part of the L-G graph offering additional information about the regions and is based on the efficient generation of the regions skeletons after the regions segmentation process. The line segments of the skeleton of a region are interrelated with each other through a graph with attributes in a similar way with the one generated by the contour of a segmented region. Thus, the skeleton graph of a region's skeleton is

$$G = K_1\ a^{c}_{12}\ K_2\ a^{c}_{23} \ldots\ldots K_k\ a^{c}_{kq} N_q \lozenge K_i\ a^{p}_{ij}\ K_j\ \lozenge \ldots\ldots \lozenge K_n a^{rd}_{nm} K_m$$

where ◊ represents the graph relationship operator, and each K_i maintains the structural features of the corresponding line segment, thus, K_i = {orientation (o), length (le), curvature (cu)}, and a_{ij} holds the relationships among these line segments, thus, a_{ij} = {connectivity (c), parallelism (p), symmetry (s), relative magnitude (rm), relative distance (rd), etc.}. The missing elements for a global visual perception of an image are: the color (or texture) of each region, its relative geographic location (distance and angle) among the other regions, its relative size in regards with the other regions, etc. One way to obtain these additional features is the development of the global image graph.

The Image Global Graph (or for simplicity the L-G graph) attempts to emulate a human-like understanding by developing global topological relationships among regions and objects. More specifically, for each image region Mi, a skeletonization task is performed and the final centroid GCg(i,x,y) is defined. When all the final centroids have been defined for every image region, the global image graph is developed:

$$\text{L-G}(Ak) = (P_1R_{12}P_2)\Phi_{23}(P_1R_{13}P_3)\ldots(P_1R_{1n-1}P_{n-1})\Phi_{n-1n}\,(P_1R_{1n}P_n)$$

where P_i is a node that represents an image region graph, its color, and its GCg(i,x,y), R_{ij} represents the relative distance between two consecutive GCg, and the orientation of each dg, Φ_{ij} represents the relative angle between consecutive distances dg(i) and dg(j). An important feature of the L-G graph is its ability to describe 3-D scenes. The only difference between 2-D from 3-D is that in 3-D the local graph will represent 3-D surfaces and the global graph will appropriately interrelate them.

3.4 Gabor filters

Wavelets are functions that are used in representing data or other functions. This idea is not new. *Approximation of information* using superposition of functions has existed since the early 1800's, when Joseph Fourier discovered that he could superpose *sines* and *cosines* to represent other functions sets. However, in wavelet analysis, the scale that we use to look at data plays a special role. Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large "window," we would notice gross features. Similarly, if we look at a signal with a small "window," we would notice small features. The result in wavelet analysis is to see both the “forest and the trees”, at the same time. They have advantages over traditional Fourier methods in analyzing physical situations when the signal contains discontinuities and sharp spikes. In particular, the most interesting dissimilarity between these two kinds of transforms (Fourier and Wavelets) is that individual wavelet functions are localized in space. Fourier sine and cosine functions are not. This localization feature, along with wavelets' localization of frequency, makes many functions and operators using wavelets "sparse" when transformed into the wavelet domain. This sparseness, in turn, results in a number of useful applications such as data compression, detecting features in images, and removing noise from time series.

One way to see the time-frequency resolution differences between the Fourier transform and the wavelet transform is to look at the basis function

coverage of the time-frequency plane. Fig. 3-5 shows a windowed Fourier transform, where the window is simply a square wave. The square wave window truncates the sine or cosine function to fit a window of a particular width. Because a single window is used for all frequencies in the Windowed Fourier Transform (WFT), the resolution of the analysis is the same at all locations in the time-frequency plane

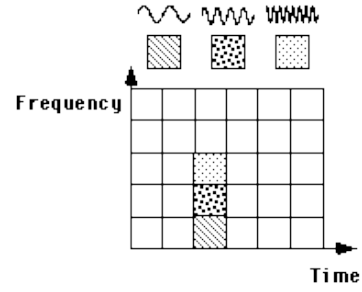


Fig. 3-5. Fourier basis functions with fixed windows in frequency/time domain.

An advantage of wavelet transforms is that the windows vary. In order to isolate signal discontinuities, one would like to have some very short basis functions. At the same time, in order to obtain detailed frequency analysis, one would like to have some very long basis functions, Fig. 3-6. A way to achieve this is to have short high-frequency basis functions and long low-frequency ones. This happy medium is exactly what you get with wavelet transforms.

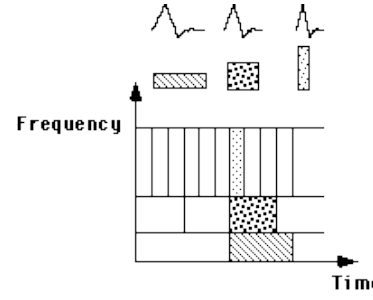


Fig. 3-6. Daubechies wavelet basis functions

There are unlimited wavelet basis functions. However, different wavelet families make different trade-offs between how compactly the basis functions are localized in space and how smooth they are, Fig. 3-7.

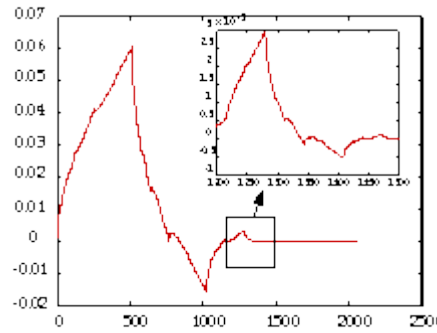


Fig. 3-7. The fractal self-similiarity of the Daubechies mother wavelet

Another wavelet family is Gabor filters that can be viewed as a sinusoidal plane of particular frequency and orientation, modulated by a Gaussian envelope.

$$h(x,y) = s(x,y)g(x,y) \quad \text{(Eq. 3-4)}$$

where the complex sinusoid s(x,y) is defined as follows:

$$s(x,y) = e^{-j2\pi(U_o x + V_o y)} \quad \text{(Eq. 3-5)}$$

And the Gaussian function is defined as follows:

$$g(x,y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} \quad \text{(Eq. 3-6)}$$

Gabor wavelets have been used in many applications, such as texture segmentation, target detection, fractal dimension management, document analysis, edge detection, retina identification, image coding and image repre-

sentation [32]. In [33] Daugmann and in [34] Webster and De Valois showed that Gabor wavelet kernels have many common properties with mammalian visual cortical cells. These properties are orientation selectivity, spatial localization and spatial frequency characterization. In this sense, Gabor filters offer the best simultaneous localization of spatial and frequency information [35].

However, while Gabor filters are very successful they suffer from a) bandwidth limitation and b) DC component. To obtain as larger spectral information while maintaining maximum spatial localization Log Gabor filters have been introduced. Log-Gabor filters have a response that is Gaussian when viewed on a logarithmic frequency scale instead of a linear one like Gabor filters. Log-Gabor filters can be constructed with arbitrary bandwidth and the bandwidth can be optimized to produce a filter with minimal spatial extent [35]. In [36] Field defines Log-Gabor filter as:

$$G(w) = e^{-\frac{[\log(\frac{w}{w0})]^2}{2[\log(\frac{w}{w0})]^2}} \qquad \text{(Eq. 3-7)}$$

where ***wo*** is the filter's centre frequency. The transfer function of the Log Gabor function is given in Fig. 3-8, below.

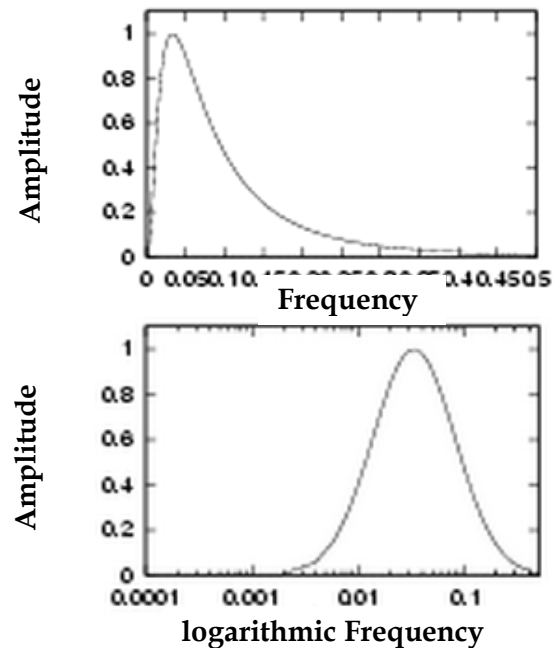


Fig. 3-8. Log Gabor Transfer functions viewed on linear and logarithmic frequency scales

CHAPTER 4: A VIDEO REGISTRATION SCHEME FOR REDUCING WCE VIDEOS

In Chapter 3 we have presented the important methods that have been utilized in our work. Here, in this chapter we will show how a novel synergistic methodology has been developed to register video frames. This registration scheme has been successfully used to register WCE video frames and therefore detect objects in consecutive frames as well as reduce the WCE video by utilizing a similarity function.

4.1 Problem Definition

The Image Registration is one the fundamental processing stages in computer vision field. It is used to geometrically align two images of the same scene taken under varying conditions. Thus, depending on these conditions, there are images taken by different sensors (multimodal registration), images acquired from different viewpoints (viewpoint registration) and images taken at different times (temporal registration). Finally, image registration can be used to match specific patterns with an image (template registration). [37]

So far, image registration algorithms have been developed to reach a very good level of efficiency and accuracy. These methods can be categorized in correlation based, frequency domain processing based and point mapping based.

Video registration, however, is still an open challenging problem due to its great complexity. There are three (3) factors that contribute to this: a) motion

and dynamics of the camera, b) objects motion and c) unknown geometry parameters (i.e. camera location, camera specifications, etc.).

There are three (3) main categories of methods for registering video frames: a) colour intensity-based methods, b) feature-based methods and c) hybrid methods [38].

As in image correlation-based registration techniques, intensity-based methods in video registration try to minimize the sum of absolute or squared differences of pixels values. For simple and small camera motions these algorithms work fine.

Feature-based methods utilize various reliable image features to find corresponding regions and perform registration.

Hybrid methods take advantage of both previous methods to perform registration.

WCE video registration is an extremely tedious task to perform because not only camera dynamics constantly change, but also because same regions in two consecutive frames change shape and characteristics [39]. To better understand this, two consecutive frames with these characteristics are shown below in Fig. 4-1 and Fig. 4-2.

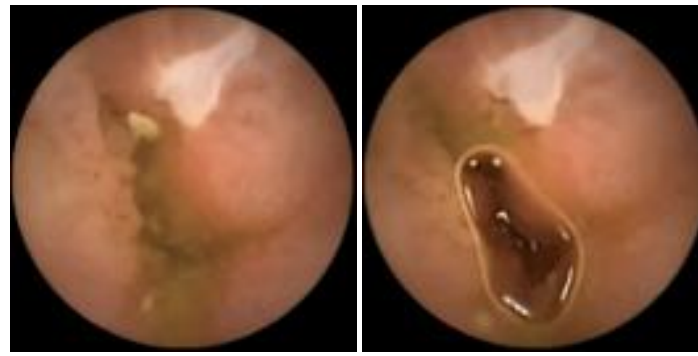


Fig. 4-1. A simple case of two successive and identical WCE video frames with shape/ characteristics dissimilarities

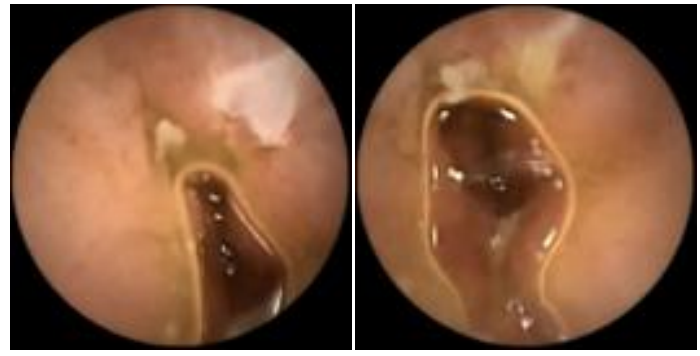


Fig. 4-2. A complicated case of two successive but similar WCE frames with shape/ characteristics dissimilarities and spatial dissimilarities

It is quite obvious that WCE video cases are very complicated and features that are to be used for the registration process need to be carefully selected. A relevant work has been done in [40], where researchers presented a methodology to reduce the WCE video frames based on clustering and non-negative Lagrangian relaxation.

Our approach to this WCE video registration problem utilized an advanced framework consisting of three (3) components: a) the colour-based features extraction phase, b) the frame structural features extraction phase and finally c) the features processing phase.

Colour-based features extraction consists of a sophisticated segmenta-

tion scheme based on work [67] in order to extract all regions of each frame.

Structural features extraction is also based on work [68] in order to provide us with information about the structure of each frame. That is the local-global graph of each frame containing connectivity information of regions.

Finally, the core of the methodology is responsible for using features extracted in the two previous steps in order to produce the registration results.

4.2 Features Processing Scheme

The steps of our registration methodology are synopsized in the following diagram Fig. 4-3

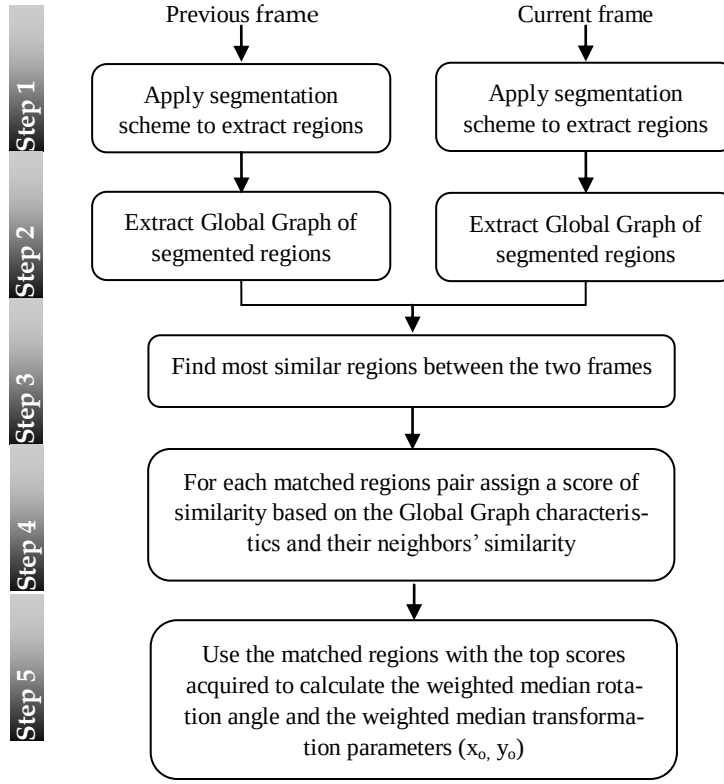


Fig. 4-3. Diagram of the proposed Registration methodology

In step 3 a repetitive process tries for all combinations of regions of both frames to find the most similar regions based on four (4) criteria:

i) Regions should not be relatively small.

ii) Regions should differ a little in size.

iii) Colour difference should be small.

iii) Regions should have very similar textures (Entropy, Contrast and Homogeneity).

In step 4 all matched pairs, identified before, are given a similarity score based on the L-G graph:

a) The orientation of each region of the matched pair is calculated with the rest of regions of the same frame and this angle is compared with each orientation of the corresponding region in the matched pair in the other frame. If the corresponding orientations match then the similarity score for this pair is increased by one (+1).

b) For each region of the matched pair, all their neighbour regions are compared using step **a**. Based on the average number of neighbours, if the majority of neighbours are similar, then the similarity score of this pair is increased by one (+1).

At the end of this process, the matched pairs are arranged according to their similarity score.

Finally, in step 5 a weighted median scheme was used to calculate the

rotation angle and the transformation parameters to perform registration, because it gave us better results than other schemes (i.e. average, mean, etc) [40].

The formula for calculating the weighted median is the following:

$$\text{(Eq. 4-1)} \qquad median = \frac{\sum_{i=1}^{n} x_i \cdot w_i}{\sum_{i=1}^{n} w_i}$$

Where x_i is the orientation between the region's centroid and the centre of the frame and w_i is the weight which in our case is the similarity score of each matched pair found in the previous step (step 4).

4.3 Degree of Similarity

Due to the unpredictable motions (i.e. bouncing) of the wireless capsule camera the probability that two successive frames are very dissimilar is relatively high. In such cases it is crucial that the methodology should not perform registration.

To support this argument we ran three (3) different WCE videos with 3125 frames, 2460 frames and 2789 frames respectively, to calculate how many dissimilar frames existed. That counts the frames that were completely different from their predecessor frames. In the first video we identified approximately 1280 out of 3125 frames, in the second video there were roughly 780 out of 2460 and in the third video there were around 1030 out of 2789. We can conclude arbitrarily that more than 1/3 of the video frames are completely dissimilar. In such cases, there is no point to perform registration.

In order to handle these cases, one could argue to compare the average color of two consecutive frames and if the color difference is above a

certain threshold then these two frames are completely dissimilar. However, this is not always a ground truth for the simplest reason that two dissimilar frames might have the same average colors. Thus, to create a more robust decision making model we implemented a quite simple confidence function of similarity which basically shows the degree of similarity of two frames:

$$\text{Sim}(m, c1, c2) = \begin{cases} 133 * m - 0.2 * \frac{c1 + c2}{2}, & \text{if } \frac{c1 + c2}{2} \leq 20 \\ 133 * m - 0.5 * \frac{c1 + c2}{2}, & \text{if } 20 < \frac{c1 + c2}{2} \leq 30 \\ 133 * m - 0.8 * \frac{c1 + c2}{2}, & \text{if } 30 < \frac{c1 + c2}{2} \end{cases} \quad \text{(Eq. 4-2)}$$

Where **m** is the percentage of matched pairs out of the total number of regions of the frames, **c1** is the distance of the weighted average colors of both frames after segmentation process (here weights are the size of the regions) and **c2** is simply the distance of the average color of both frames.

The design of (Eq. 4-2) was carefully planted after numerous experiments with videos and cautious examination. For example, the coefficient of m was assigned value 133 after having run a total number of 20 artificially-made videos. Each of these videos was carrying copies of a specific WCE image which was randomly chosen from a total number of 20 WCE images. We noticed that **m** was reaching an average value of 0.75. One would except that the average value should be 1.00 since the frames are exact copies. However, this difference is caused by: i) the randomness and fuzziness of the segmentation process, ii) the level of detail of segmentation was not set at the highest and iii) the strict criteria of step 3, previously described.

For c1 and c2 we know that there is a linear relationship among them.

c1 is more robust because it takes into account not only the colour of the regions but also the size of the regions. We decided to include c2 for the reason that it is also a reliable measure, and since c1 is usually bigger than c2, c2 tends to pull the average value lower, which is true. The choice of the coefficients 0.2, 0.5 and 0.8 was made after many experiments.

We recorded the values of m, c1 and the output of similarity function from 70 frames. The results are presented in the graphs below (Fig. 4-4, Fig. 4-5).

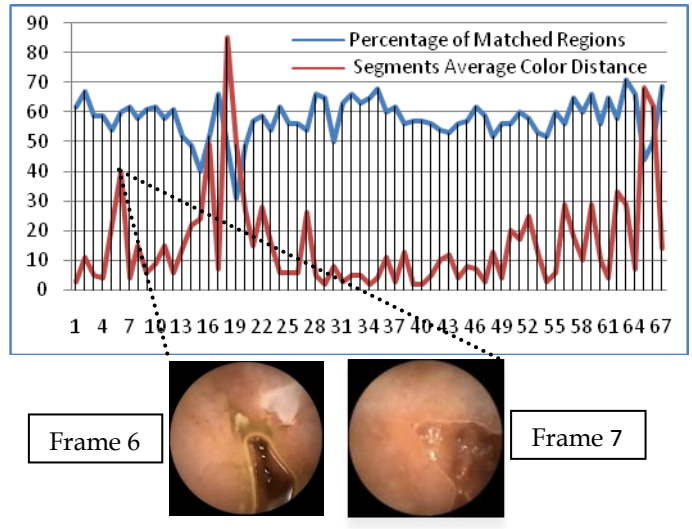


Fig. 4-4. Percentage of Matched Regions (see parameter m) with blue color and Segments Average Color Distance (see parameter c1) with red color for a 70-frame WCE video

Some useful conclusions can be reached after examining the above graphs.

In Fig. 4-4 the importance of using, parameter m in the similarity function is quite obvious. For example, a quick overview of frames 6 and 7 shows

that these frames are identical though their color difference is relatively high (m = 54, c1= 22, c2=16 and Similarity = 68%).

In Fig. 4-5 similarity is averagely 75% showing that all frames, which are near this value, are similar. When similarity is 60% - 65%, frames are relatively dissimilar but the registration scheme can be run. However, when similarity is below 60% then the frames are very dissimilar, like the ones shown in Fig. 4-5 (frames 18 -19 and frames 65 -66).

Further investigation is under process to examine the utilization of more than 2 frames. In this way more robust results may be achieved since there is a higher level of observance.

Having integrated the similarity function into the rest of the methodology, the algorithm runs step1, step 2 and step 3 and then checks the similarity of the frames. If they are quite similar (>65%), it continues executing steps 4 and 5, otherwise it reports a new series of frames and moves on.

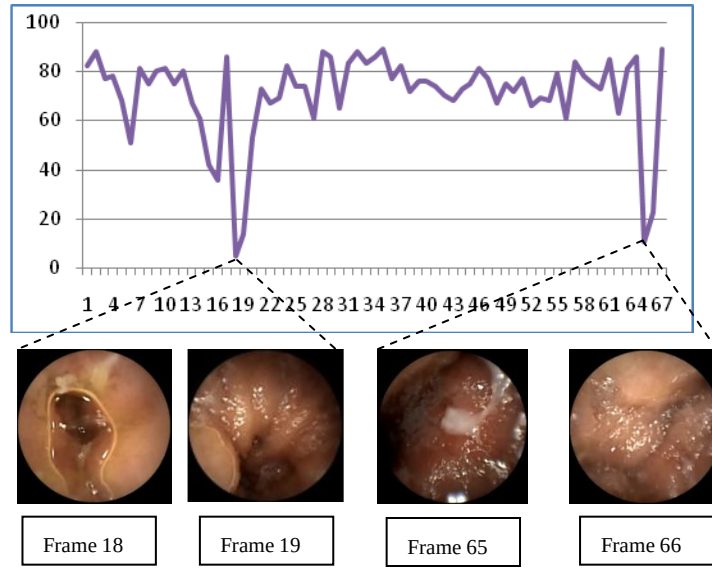


Fig. 4-5. Similarity function output for the same 70-frame WCE video

4.4 The Center of Image Similarity

Our methodology was implemented in the ATRC Video Analysis toolbox that runs under GUI platform and offers many capabilities for better user friendliness. Because our end users who test our toolbox are medical doctors, we created the feature of the Center of Image Similarity in order to offer a better spatial comprehension of the similarities in WCE frames.

Simply enough, the Center of Image Similarity creates a grid of squares (i.e. 4x4, 8x8 etc.) on the current frame. Each square is painted grayscale if there are any top similar matched regions in that squared area. Grayscale value is selected by the percentage of area the top similar regions cover in the specific box. A graphical representation of the Center of Image Similarity is shown below (Fig. 4-6).

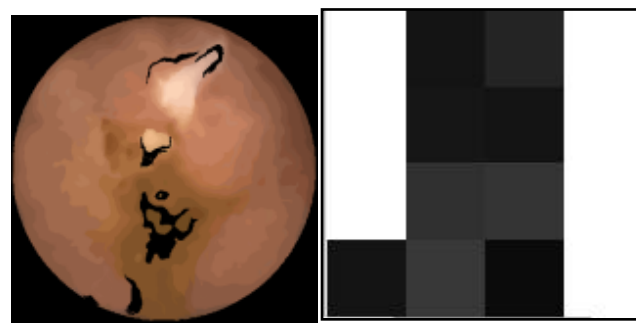


Fig. 4-6. Frame with top 9 similar regions (black painted) are shown on the left. Corresponding Center of Image Similarity is shown on the right. Boxes with higher level of gray have higher similarity than boxes with lower gray level. White boxes carry no information.

Physicians at Digestive Associates in Dayton, OH have welcome the use of this feature because as they have now a better understanding on which parts of two consecutive frames are highly similar.

4.5 Experimental Results

We ran our methodology with various WCE videos. In this section we present snapshots of a certain WCE video taken at various times, Fig. 4-7.

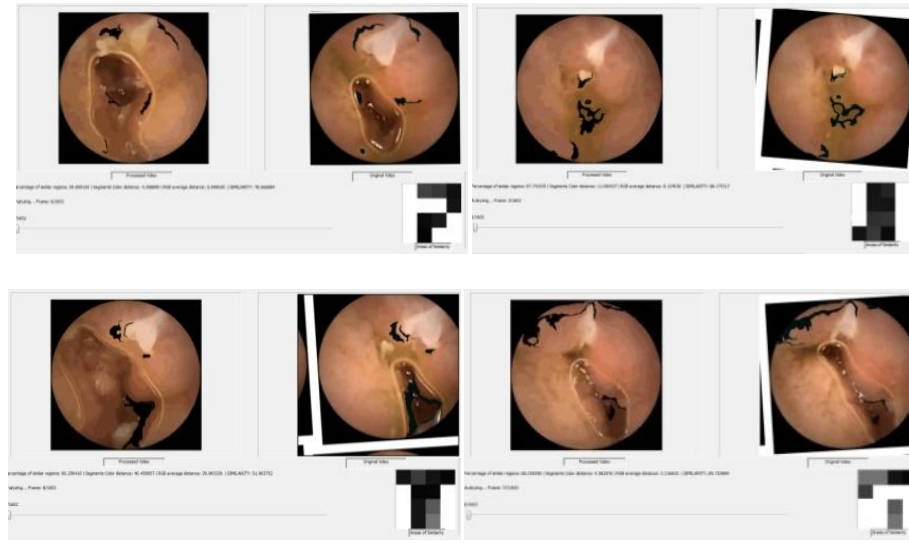


Fig. 4-7. Four cases where registration scheme is applied. The left frame corresponds to the current frame, whereas the right one corresponds to the previous frame. Black painted regions are the ones that are identified as the top matched regions

The methodology here attempts to identify corresponding regions in two successive frames very accurately. For example, although in the last screenshot the two frames carry the same information, they are quite dissimilar. The registration scheme identified the right regions and achieved high accuracy in the registration.

We further evaluated our registration scheme in big scale experiments. We used three (3) different real patient videos. Out of each video we randomly extracted a 750-frame sequence. The purpose of the experiments was dual: a) how successful the frame reduction was and b) how successfully the registra-

tion scheme performed between similar frames.

We performed these experiments in collaboration with our associate physician Dr. Pouagare and his team at Digestive Associates. We handed them out the three (3) videos and asked them to review them and set timestamps at those frames they believed that a new series of frames started. We took their answers as the ground truth. We performed the same procedure using our proposed algorithm. The results are shown in Table 4-1 below.

	ATRC algorithm	**Digestive Associates team**
Video 1	279	127
Video 2	365	226
Video 3	290	184

Table 4-1. Frames Reduction results

It can be seen from Table 4-1 that our algorithm managed to reduce the frames down to an average of 32% of the 750-frame videos, whereas the physicians reduced the videos to 24%. Our algorithm performed very well. However, what really matters is how many frames out of our methodology-based identified frames were identical with frames that the physicians identified as important, which is taken as the ground truth.

We also calculated sensitivity and specificity of our methodology on the three (3) videos.

	Important Frames	Non Important Frames
Identified	116	163
Not Identified	11	460

Table 4-2. Video 1

	Important Frames	Non Important Frames
Identified	209	156
Not Identified	17	368

Table 4-3. Video 2

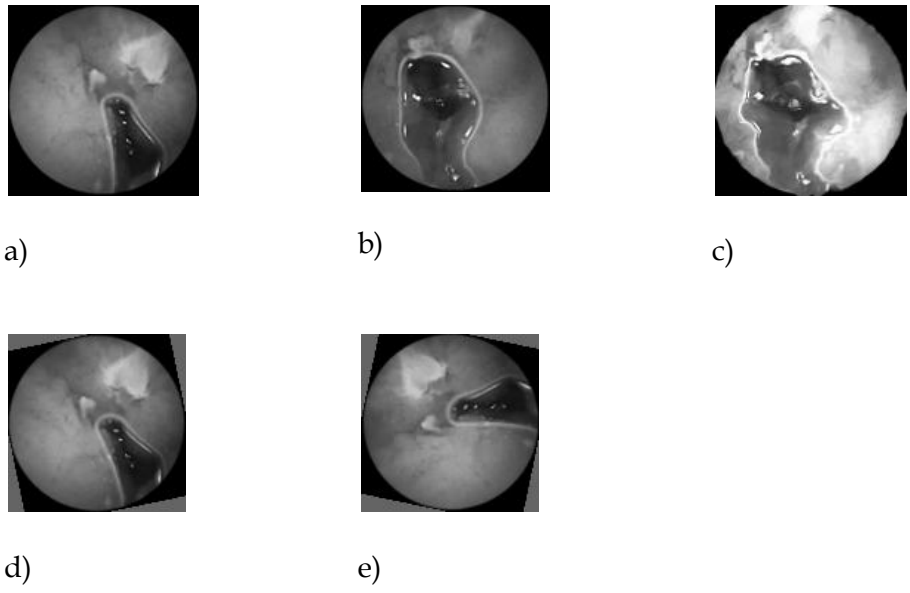
	Important Frames	Non Important Frames
Identified	168	122
Not Identified	16	306

Table 4-4. Video 3

	Sensitivity	Specificity
Video 1	91%	73%
Video 2	92%	70%
Video 3	91%	71%
Average	91%	71%

Table 4-5. Sensitivity and Specificity for the three (3) videos

We performed a registration comparison with other widely-used registration methodologies. For this reason we run our dataset with three (3) popular algorithms of ITK toolbox. In Fig. 4-8 below you can see the results for various registration methodologies.



a) b) c)

d) e)

Fig. 4-8. a) Fixed image, b) Moving image, c)FEM registration, d)Rigid registration in 2D, e) Registration with Centered Similarity 2D Transform

In order to get quantitative comparison results between our methodology and the three (3) methodologies we used the ratio of Intensity(fixed image)/Intensity(transformed image) for the success of the registration. For a perfect successful registration the ratio is 1. We run all the algorithms with the same WCE datasets and the results are synopsized in the following table:

	Average similarity ratio
LG Registration	87.56
FEM deformation	79.01
2D Rigid registration	81.94
Registration with Centered Similarity	84.23

Table 4-6. Comparison of registration methodologies

4-6. Conclusion

In this chapter a novel approach for the video registration problem was

presented. The methodology was applied to the registration of WCE frames and proved to behave very effectively. Since this methodology works in difficult situations, like WCE videos, it can be applicable to other registration cases too. Although the algorithm has reached a high degree of sensitivity, its specificity is not at the same level. This can be easily explained due to the high number of non important frames categorized as important ones. It must be clarified that "Important Frames" refer to the frames identified by physicians, whereas "Non Important Frames" refer to the frames not identified by the physicians. Since the degree of similarity of the methodology is quite sensitive to changes between two frames, it is normal that trivial frames are identified as important ones, therefore decreasing specificity.

Besides the frame reduction evaluation, we tested frame registration on the three (3) videos. However it proved to be quite difficult to quantify registration success due to the ambiguity of the WCE videos, which can be witnessed at the illustrative examples above. In order to overcome this issue and to reach objective conclusions we performed the following tasks:

a) In each video the algorithm identified a certain number of unique frames as starts of series of same-content frames (Table 4-1). We used these frames to create groups of similar frames. For example, in video 1, unique frame 184/279 has similar frames the next five (5) frames.

b) After registration process these frame groups were handed to three (3) members of ATRC team. Each member was asked to evaluate registration results on each couple of frames in each group in all three (3) videos.

c) Registration success rates were given in a scale of five (5) levels: "Registration Failed", "Registration Unsatisfactory", "Registration Average", "Registration Satisfactory" and "Registration Excellent". The discussion between members on setting standards for each level of registration success is in progress. After presenting various cases of registration results there was a common basis. The results of our discussion with the ATRC team and the Digestive Associates team will be announced.

Further research is carried towards the improvement of the methodology. We think that the weighted median scheme used here to calculate the registration parameters (rotation and transformation) should be replaced with a more advanced scheme.

Additionally, we will try to extend the methodology not to process just two frames at the time but five frames which will add higher accuracy and more robustness in the degree of similarity.

Moreover, our goal is to create a sophisticated L-G graph scheme that will offer detailed description of a frame's structure in order not only to improve the registration methodology, but to provide to us an intelligent or ad-

vanced description of the differences between two frames.

By increasing the level of detail of segmentation, processing time increases but the accuracy of the registration improves dramatically. This showed to us that a hardware implementation of the scheme will lead to a faster and more accurate video registration, with a wider potential use in many other areas (i.e. aerial photographing under tough circumstances).

CHAPTER 5: DETECTION OF BLOOD-BASED ABNORMALITIES IN WIRELESS CAPSULE ENDOSCOPY VIDEOS

5.1 Introduction

Although RAPID READER from GivenImaging offers a blood detector as discussed in chapter 1 its performance is debatable. A work in 2003 [42] assigned 72% sensitivity and 85% specificity to the blood indicator while a work in 2005 [43] gave 37% sensitivity and 59% specificity. This leaves the detection of bleeding regions in the small intestine still an open and crucial issue. Moreover, a good detection classification approach of the blood based abnormalities in the intestine is missing by the techniques used in this research area according to the gastroenterologists. Thus, this chapter offers a novel approach to automatically detect and categorize blood based abnormalities in the intestine by starting with their definitions.

Blood-based abnormalities in the small bowel are characterized in three categories: a) Bleeding, b) Angioectasia and c) Erythema. Such a categorization is important in order to comprehend the problem efficiently.

A) Bleeding

Bleeding is defined as the flow of blood from a ruptured blood vessel into the digest tract. Among malignant tumors, leiomyosarcoma is most commonly associated with bleeding. Bleeding with adenocarcinoma is less frequent, and rarely occurs with carcinoid tumors [44].

An inside view of the digestion tract is necessary to reveal possible bleedings. This can be feasible by using the WCE technology, which is a non-invasive and painless procedure compared to conventional endoscopic procedures like gastrointestinal procedure and enteroscopy. In a multicenter research by Pennazio et al, [45], 29 of the 60 patients underwent push enteroscopy in addition to capsule endoscopy. Among those 29 patients, push enteroscopy detected a source of bleeding in 8 patients (28%), whereas capsule endosopy detected a source of bleeding in 17 patients (59%).

B) Angioectasia

In medical bibliography angioectasia is also referred to as arteriovenous malformations. It is the most common abnormality accounting for obscure gastrointestinal bleeding, seen in 21%–53% of patients who undergo capsule endoscopy [46]. This makes their detection an important task. They occur more frequently with increasing age and can be identified at endoscopy as spiderlike lesions [47]. Since most of angioectasia cases in small intestine are caused by blood vessels inside the intestine walls. In comparison with bleeding cases, angioectasia stays inside the intestine walls. The color of an angioectasia appears often more reddish than the color of a bleeding.

C) Erythema

In general, erythema is defined as skin redness caused by capillary congestion. In the small intestine erythema multiforme is strongly related with

various abnormalities (i.e. Crohn's disease). It is an acute, self-limiting, inflammatory skin eruption. Its redness is lesser than that of angioectasia.

An overview of the three cases is presented below (Fig. 5-1).

Fig. 5-1. Blood-based abnormalities: Bleeding (left), Erythema (middle), Angioectasia (right)

5.2 The Blood Based Detection Scheme

The methodology proposed here is based on a synergy of techniques for accomplishing its results, Fig. 5-2.

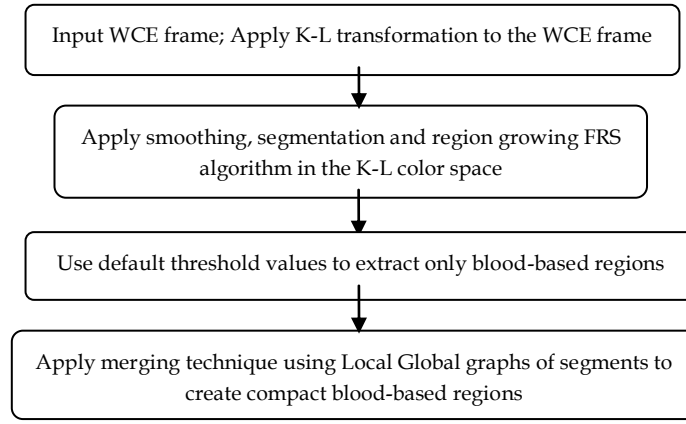


Fig. 5-2. The synopsis of the Detection methodology

5.2.2 Color Transformation Scheme

One could argue that since bleeding regions in WCE images have dominant R component in the RGB color space, their identification must be an easy task. However, this is not always true. After a quick examination, one could easily see that most WCE image pixels carry high values of R, thus dismissing this assumption. It is necessary that other methods must be utilized to approach this problem.

We know that R, G, and B components are highly correlated Fig. 5-3. We need to make the input feature space as much uncorrelated as possible, so that more useful information can be extracted. That's the reason why in image processing area other color spaces are widely tried and used.

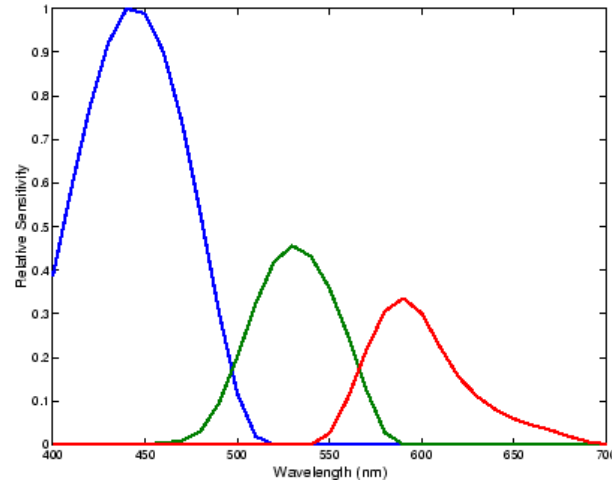


Fig. 5-3. CCD Color Spectral responses for Red, Green and Blue Component

Keeping in mind that RGB color space doesn't offer the visualization ability to enhance and make bleeding regions distinguishable, we tried some major color spaces like HSV, Lab, Luv, YDbDr and CIE-XYZ. YDbDr produced

some interesting results but even those didn't prove to be satisfactory. (Fig. 5-4)

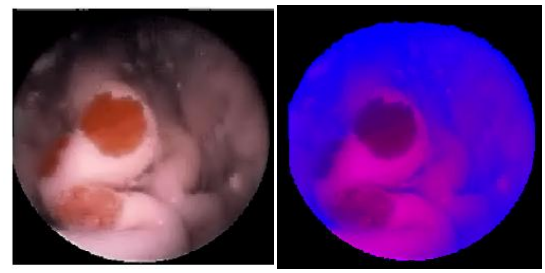


Fig. 5-4. Original WCE image (left) in RGB color space and in YDbDr color space (right)

Thus, a solution comes from the work of Yu-Ichi Ohta et al, [50]. In their paper, the authors produced a set of color features using Karhunen Loeve transformation of R, G, B components after having evaluated 8 different types of natural images.

The usefulness of a color feature is greatly influenced by the structure of color scenes to be segmented. This statement forced us to investigate the Karhunen-Loeve transformation in our case.

It is well known that R, G and B components overlap in the frequency domain (Fig. 5-3) and so being highly correlated. Therefore, the need for a transform that can make the R, G, B vectors as much uncorrelated as possible is obvious. We can remove correlations between pixels using an orthogonal linear transform called Karhunen Loeve transformation [50].

The choice of a linear transformation like Karhunen-Loeve is intentional. As mentioned in the work of Kender [52], linear color space transformations are more preferable than non-linear ones for the following reasons:

1) non-linear transforms suffer from non-removable singularities

2) their values are distributed in such way creating spurious modes and gaps and

3) non-linear transforms have higher computational complexity.

On curved surfaces the intensity often gradually changes and cannot be used as a useful feature for segmentation. Color information in natural scenes is almost two dimensional (intensity and one chromatic feature).

In pattern recognition a feature is said to have large discriminant power if its variance is large. Thus, we tried to derive color features (color space) with large discriminant power. More specifically, as described in [50], let S be the region to be segmented and Σ be the covariance matrix of the distributions of R, G and B in S. Let λ1, λ2 and λ3 be the eigenvalues of Σ and $\lambda 1 \geq \lambda 2 \geq \lambda 3$. Let $Wi = (W_{Ri} W_{Gi} W_{Bi})t$ for i=1,2,3 be the eigenvectors of Σ corresponding to λi, respectively. The color features X_1, X_2, X_3 are defined as:

(Eq. 5-1) $Xi = W_{Ri} \cdot R + W_{Gi} \cdot G + W_{Bi} \cdot B$ $(\|Wi\| = 1, i=1, 2 \text{ and } 3)$

It is well known in mathematics that the derived X1, X2 and X3 vectors are uncorrelated.

After many experiments the authors reached to the conclusion that W_1 is dominated by $(1/3 \ \ 1/3 \ \ 1/3)^t$, W_2 is dominated by $(1/2 \ \ 0 \ \ -1/2)^t$ or $(-1/2 \ \ 0 \ \ 1/2)t$ and W_3 is dominated by $(-1/4 \ \ 1/2 \ \ -1/4)^t$.

Note that from our experiments we have concluded that $W_2 = (1 \ 0 \ -1)^t$ gave us better discriminant power.

5.2.3 *Region Synthesis*

The synthesis of regions is an important step that leads to patterns or objects recognition. The Local-Global graph method used here can match a pattern or an object with deformation other than rigid transformation. However, it assumes that the shape has been segmented from the background, and the mathematical shape representation is sensitive to some kinds of deformations, for example, if the shape of a region changes the local graph and the skeleton graph record the changes attached them to the L-G graph. In addition, the centroid may change and the global graph will register that change as well. The process used here for objects recognition is based on the synthesis of segmented adjacent regions, using the L-G graph, and association (comparison) of the integrated region models available in a database. Fig. 5-5 shows the synthesis of the regions based on the L-G graph in 5 steps for the generation of a new region that the physician will consider it as one. Note that the final region regains the RGB values of its pixels.

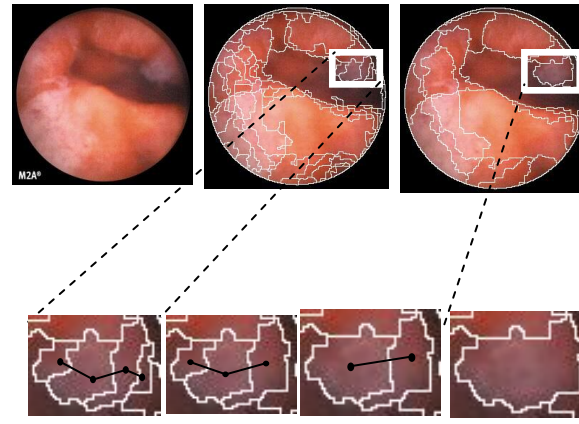


Fig. 5-5. Original image and Segmentation with Region Merging / Synthesis

5.3 Experimental Results

We have tested our methodology using images available in the RAPID Atlas ® which is a library of reference images, offered by Given Imaging ©.

Some identified blood based regions frames are shown in Fig. 5-6, Fig. 5-7, and Fig. 5-8. The algorithm offers the option to give the number of blood based pixels as well as the percentage of blood in the given frame, so that the physician has better understanding of the seriousness of the case.

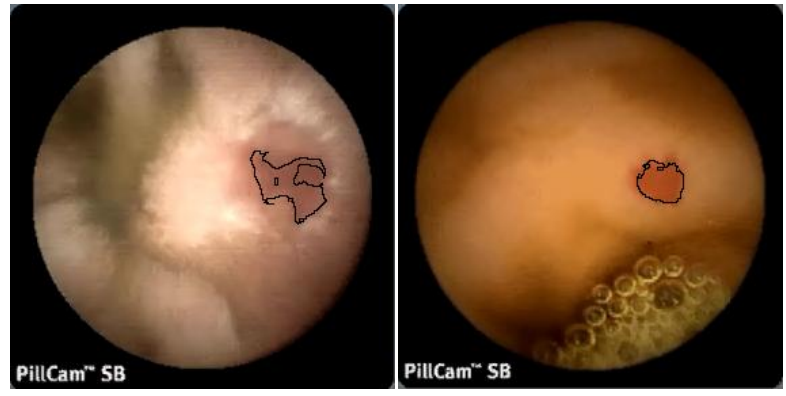


Fig. 5-6. Angioectasia

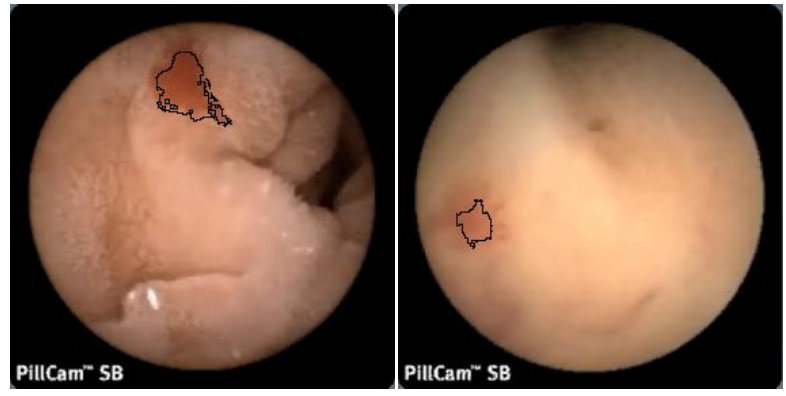


Fig. 5-7. Erythema

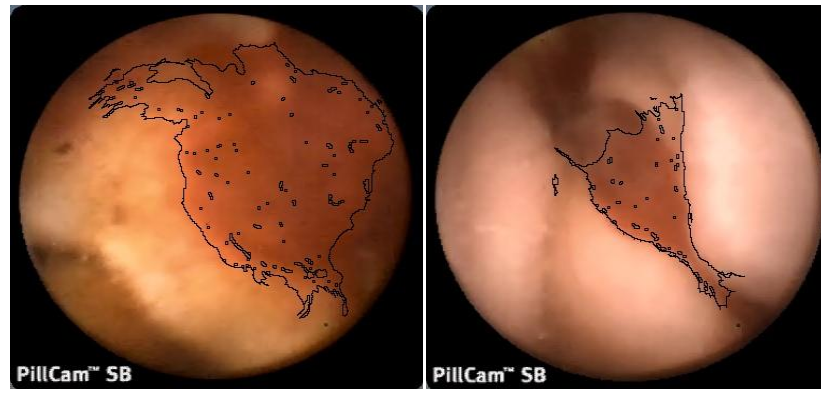


Fig. 5-8. Bleeding

It is important to be mentioned that the methodology should identify regions that may be suspicious even though after physician's examination, are proved not to be. This is a desirable feature that gastroenterologists want present in any automated software methodology.

We collaborated with physicians from the Digestive Associate in Dayton, OH and they provided us with capsule endoscopy video cases. The dataset consisted of 55 frames containing blood-based abnormalities and 160 frames of normal tissue. It must be mentioned that the collection of such a dataset requires long time because it is very rare that a patient can have multiple cases of blood-based abnormalities. The physicians point out the right the blood-based abnormalities. The methodology was run on the dataset and produced the following results:

	Blood-based abnormalities	**Non blood-based abnormalities**
True	50	31
False	5	129

Table 5-1. Blood-based methodology: Sensitivity: 90.90% Specificity: 80.65%

To better evaluate our methodology on the detection of blood-based abnormalities we performed an elaborate search on the scientific and engineering literature. We found out that by February 2010 there have been six (6) major research works that deal exclusively with the detection of blood-based abnormalities.

In [48], the authors confused clotted blood with bleeding. Although their technique aims to detect bleeding regions, it eventually detects blood regions. In their work, the 2D color distributions show a significant overlapping of non-blood and blood pixels, though they claim the opposite. In addition, as stated above, a simple examination of the bleeding pixels and non-bleeding pixels can prove that their values are identical in many cases. Dismissing dark pixels from the image doesn't necessary imply removal of non-blood pixels. There are cases when the illumination of the WCE is low making bleeding regions darker. Their methodology dismisses small detected blood regions. It is needless to mention of the existence of small-sized bleeding regions. However, the use of Expectation Maximization clustering and Bayesian Information Criterion are very promising tools, offering an automatic clustering and a statistical approach to this difficult problem.

In [49] HSV color space is used to enhance WCE image so that bleeding regions can be extracted easier. It is true that bleeding areas appear better with high color saturation but as the authors state themselves most of the GI tract presents a red dominant color and for highly illuminated GI tract images; the saturation may take very high values making distinction between bleeding and

non-bleeding regions impossible. Although their methodology is supposed to consist of multiple features, only saturation and intensity of color are used on step 1 and step 2, which can be further unified into one. In addition the use of multiple and numerous empirical thresholds in step 1 and step 2 makes the methodology less reliable and robust.

In [124] authors utilize the Reed-Xiaoli (RX) detector [125] as the core of their bleeding detection methodology. RX detector is used to discriminate the bleeding regions from the surrounding normal tissues based on the typical formulation assumes that the covariance matrix of the data is unknown, and that anomalous pixels (bleeding pixels) differ from the background in their mean value. Again in this paper the importance of discriminating the bleeding pixels from the rest of the normal pixels is raised. This is shown by the authors when mentioning that the gastrointestinal images present high red hue values. Whereas our proposed methodology takes advantage of covariance concepts straight on the color, their methodology takes advantage of the covariance matrix to discriminate bleeding pixels from a statistical perspective.

In [126] the authors use a color spectrum transformation to detect the active bleeding regions. Similarly to [47] and [49] the paper lacks of sufficient results to prove their concept. Additionally the use of multiple consecutive empirical thresholds does not increase robustness. The use of a color transformation is not explained well so that the reader can understand its importance. However, the use of brightness adjustment pre-processing techniques looked very promising.

In [127] the authors used an SVM ensemble to create a classification method for finding video frames that contain bleeding lesions. The input to the SVM classifier were a) the histograms of hue, saturation, and value, b) dominant color descriptor and finally c) the co-occurrence matrix of the dominant colors. The main focus of their research has been on the training and optimization of the SVM ensemble and less on researching on the selection and processing of optimal features. Additionally, their goal was not to identify the bleeding regions themselves but the potential abnormal video frames. However credit must be given for this serious work and especially on the classification side.

In [128] the authors present a methodology for detecting bleeding regions in WCE frames by utilizing Tchebichef Polynomials to describe color moments. The output is used to produce texture features with Local Binary Pattern texture operator. Finally the texture features are fed into a traditional multilayer perceptron neural network for classification. Although the methodology is straightforward the results promising the authors do not clarify and they do not point out how the color moments help them with better discriminative power on bleeding pixels.

Since it is impossible to test all the above methodologies on the same data we compared the detection rates of each methodology as mentioned in every paper. We used the sensitivity and specificity rates and we came up with the following Table 5-2.

	Sensitivity	Specificity
Expectation maximization clustering [48]*	0.9255	0.9810
Multiple Features [49]	0.8833	0.6707
Reed-Xiaoli (RX) detector [124]*	0.9200	0.8800
Color spectrum transformation [126]	0.9286	0.8949
SVM classification [127]*	0.8116	0.9334
Color moments [128]*	0.9113	0.9313
K-L based transformation (proposed method)	0.9090	0.8065

Table 5-2. Comparison table for various methods of bleeding detection

*(The * symbol indicates insufficient information about their data sets.)*

The methodology in [48] showed impressive results. However as mentioned above there seems to be a confusion of clotted blood with bleeding pixels.

We believe that the specificity rate of our methodology can be increased if we utilize the knowledge of the three (3) types of blood-based abnormalities as discussed in 5.1.

After an evaluation of the Table 5-2 results and the discussion above, we can conclude that the use of the magic triplet: a) discriminative color/pixel transformation (Reed-Xiaoli (RX) detector [124], Color spectrum transformation [126], K-L transformation), b) texture features (Color moments [128]) and c) classifiers (SVM classification [127], Color moments [128]) can lead to a sophisticated methodology that can deal with the difficult problem of the identifica-

tion of blood-based abnormalities.

5.4 Conclusions

In this chapter a new synergistic methodology was proposed for extracting bleeding regions in WCE frames based on the orthogonal Karhunen – Loeve transform.

The Local Global (L-G) graph scheme proved to be a powerful tool in our methodology since it offered us with the power of geometry description of extracted blood-based regions. Having this "description" in our hands we were able to merge smaller regions into bigger solid ones using synthesis method. In addition the L-G graph scheme is enabling us to use the geometric and color characteristics of the regions in order to categorize the abnormalities more efficiently.

CHAPTER 6: IDENTIFICATION OF POLYPS

In this chapter our ideas and research directions towards the detection of the important intestinal abnormality of polyps are given. We are briefly discussing our approach and thoughts for the detection problems. The synergistic methodology, the research and the experiments are also provided here.

6.1 What are Polyps?

Polyps are growing tissues inside the human body. Although polyps usually appear in the colon, stomach, and urinary bladder they may also appear in the small intestine. A slice of a polyp is shown in Fig. 6-1, below.

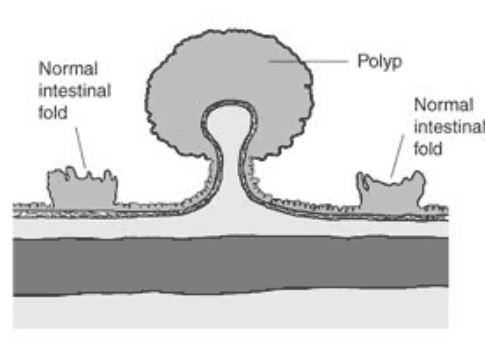


Fig. 6-1. Graphic representation of a polyp in human colon. Courtesy of National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health

Polyps can either be benign or non-benign. Most polyps are not cancerous. However, since they can turn into cancer, physicians remove them and test them by performing biopsy [53]. Sometimes polyps can bleed causing anemia, while if they are bigger than 1 centimeter have a greater cancer risk

associated with them than polyps under 1 centimeter.

Before the appearance of wireless capsule endoscopy it was impossible for medical doctors to examine the small intestine for polyps non-invasively. A more invasive procedure has been video gastroscopy which helps physicians examine the upper gastrointestinal tract (esophagus, stomach and duodenum).

Although the use of gastroscopy is limited to the upper digestive tract, it allows doctors perform biopsies and/or even resection of polyps, whereas WCE technology cannot do so or at least for now.

Besides WCE and gastroscopy technologies, computed tomography colonography (CTC) or virtual endoscopy is used mainly for detecting polyps in the colon non-invasively. Due to the nature of the virtual endoscopy technology, 3D reconstruction of the colon is feasible. In Fig. 6-2 a slice of the colon capture by CTC is given.

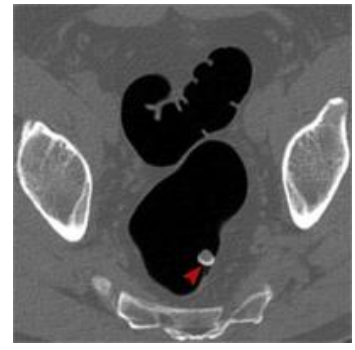
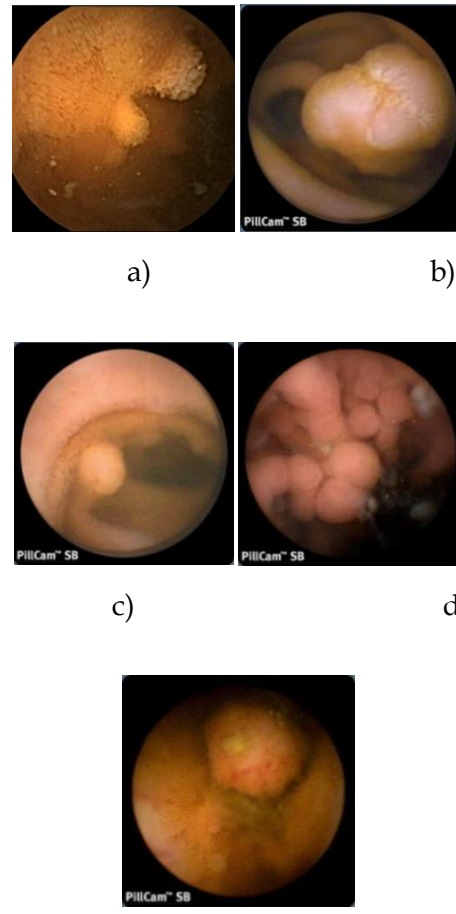


Fig. 6-2. Polyp pointed with arrow in virtual endoscopy image. American Journal of Roentgenology

One can see the resemblance between Fig. 6-1 and Fig. 6-2. In CTC polyps can be identified since they protrude into the colon lumen (black colour in Fig. 6-2). Although CTC is very promising for cancer detection in the colon, this

is not the case for the small bowel. Due to limitations of this technology, CTC is not capable of identifying polyps smaller than 1cm in diameter making it almost pointless to be used in the small intestine. Additionally, radiation exposure is a downside of virtual endoscopy.

On the other hand, WCE is the least invasive and least harmful ensdoscopy technology with few side effects. Although it offers an inside view of the digestive tract polyps have great shape irregularity, as it can be seen from the five cases in Fig. 6-3, below.



a) b)

c) d)

e)

Fig. 6-3. Polyps in WCE video

Due this variety in shape, colour, texture and size of polyps in WCE videos detection is extremely hard. Polyps appear elliptical (Fig. 6-3-b, e), round (Fig. 6-3-c, d), semicircle (Fig. 6-3-a) and in different sizes (Fig. 6-3-a, b). Although human perception makes it easy to identify polyps, computers aided software must utilize sophisticated techniques and method for automatic detection of polyps in WCE.

Some research has been done in automatic detection of polyps in WCE videos. In [54] authors use Discrete Wavelet Frame Transform (DWFT) to extract texture features (Color Wavelet Covariance). In [55] authors propose a simple methodology that utilizes color and position features to identify polyps.

However, significant research has been done on automatic detection of polyps in CTC. In [56] researchers use shape index, curvedness and sphericity ratio as geometric features for polyp candidates. In [57] and [58], authors use curvature information to extract regions of high curvature and therefore polyp candidates. In [59] authors propose geometric characteristics of polyp shapes: polyp height, polyp radius, polyp boundary length and polyp mean intensity.

6.2 Issues with the detection of polyps

It is worth mentioning at this point that polyps' detection in CTC is a very difficult task and all the above CTC methodologies are complex and sophisticated in order to reach a high level of accuracy. However, suspicious regions can be also found on the boundary of the colon lumen (Fig. 6-2), while in WCE videos polyps can be found anywhere (Fig. 6-3). Extraction of polyp can-

didates in WCE cannot be based on the lumen boundary because in most cases it cannot be seen due to camera's perspective. Therefore in case of polyp identification in WCE the most crucial part is the pre-processing step which is the segmentation scheme. The segmentation process must fulfill two goals:

a) Maintain details of object boundaries.

b) Extract only crisp segments.

However, these two goals cannot be successfully satisfied at the same time using traditional segmentation algorithms based on colour or texture values, because these algorithms can either focus on high detail producing many segments or focus on crisp segments losing boundary details. Therefore, in order to overcome this issue, we need to search for a different approach of segmentation.

6.3 The Synergistic Methodology

As described in 3.4 section Log Gabor filters proved to be ideal wavelets to perform segmentation on WCE images. Truly, Log Gabor filter banks successfully managed to extract crisp regions in WCE video frames (Fig. 6-4).

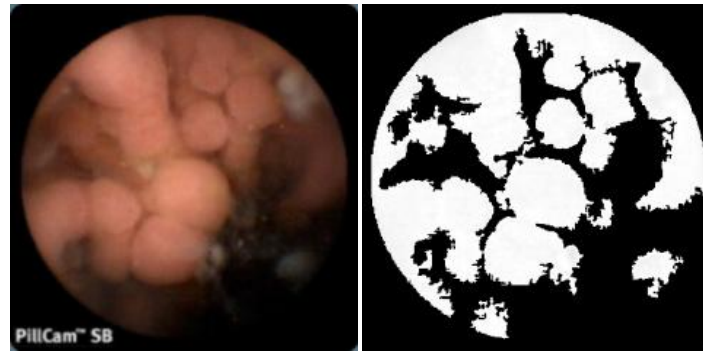


Fig. 6-4. WCE video frame containing polyps and corresponding log Gabor filter output

As it can be seen from Fig. 6-4 log Gabor segmentation process creates meaningful binary regions following at a great extent the human visual system as well-pointed by field.

6.4 The Susan Edge Detector

To further increase the robustness of our methodology, we increased the level of detail of the boundaries by using the promising SUSAN edge detection scheme.

SUSAN stands for Smallest Univalue Segment Assimilating Nucleus and it is described thoroughly in [73]. It follows the traditional method of taking an image and using a predetermined window centered on each pixel in the image it applies a locally acting set of rules to give an edge response. This response is then processed giving as the output a set of edges.

More specifically, a circular mask is placed at each point in the image and, for each point; the brightness of each pixel within the mask is compared with that of the nucleus (the centre point):

$$c(\vec{r},r\vec{o}) = \begin{cases} 1 \, if \mid I(\vec{r}) - I(r\vec{o}) \mid \leq t \\ 0 \, if \mid I(\vec{r}) - I(r\vec{o}) \mid > t \end{cases} \quad \text{(Eq. 6-1)}$$

where $r\vec{o}$ is the position of the nucleus in the two dimensional image, $\vec{r}$ is the position of any other point within the mask, $I(\vec{r})$ is the brightness of any pixel, t is the brightness difference threshold and c is the output of the comparison. However using a mask with Gaussian weighting it gives

(Eq. 6-2) $$c(\vec{r},\vec{ro}) = e^{-(\frac{I(\vec{r})-I(\vec{ro})}{t})^6}$$

a smoother version. This allows a pixel's brightness to vary slightly without having too large an effect on c, even if it is near the threshold position. The use of the sixth power can be shown to be the theoretical optimum. Summing the results for each image pixel within the mask gives us:

(Eq. 6-3) $$n(\vec{ro}) = \sum_{\vec{r}} c(\vec{r},\vec{ro})$$

This is the number of pixels inside USAN (Univalue Segment Assimilating Nucleus). Next the response of the edge is calculated by equation:

(Eq. 6-4) $$R(\vec{ro}) = \begin{cases} g - n(\vec{ro}) & if \quad n(\vec{ro}) < g \\ 0 & otherwise \end{cases}$$

where g is the geometric threshold of SUSAN algorithm and is set to 3 * nmax / 4 , nmax is the maximum value which n can take (max kernel area) and it is necessary for images that suffer from noise. From last equation it can be concluded that the smaller the USAN, area, the larger the edge response. In other words this means that the larger the USAN area, the more probable the area under examination is a surface rather than an edge. In Fig. 6-5 the output of SUSAN edge detector for a WCE video frame is given.

Fig. 6-5. WCE video frame containing polyps and corresponding SUSAN edge detection output. Brightness_Threshold=5, USAN_kernel_radius=5

6.5 Curvature center points & Geometric Rules

Using SUSAN edge detection output and log Gabor output we manage to produce crisp segments with very detailed boundaries. However, these segments as seen in Fig. 6-4 are still merged into one big region and therefore they need to turn into single meaningful regions.

To do so, the algorithm runs along the boundaries of the binary image Fig. 6-4 calculating the curvature of each point. For these points that have certain curvature and above the centers of curvature are found and stored.

The center of curvature (ξ, η) for a boundary pixel (x,y) is given by the following formula:

$$\xi = x - \frac{(x'^2 + y'^2)y'}{x'y'' - x''y'}, \quad \eta = y + \frac{(x'^2 + y'^2)x'}{x'y'' - x''y'}.$$

(Eq. 6-5)

Where $x' = f(i,j+1) - f(i,j)$, $x'' = f(i,j-1) - 2 * f(i,j) + f(i,j+1)$, $y' = f(i+1,j) - f(i,j)$ and $y'' = f(i-1,j) - 2 * f(i,j) + f(i+1,j)$, f() is the intensity of the image and i,j the coordinates of the pixel.

This results two clouds of curvature center points inside the white segments Fig. 6-4. Still, in order to create single regions a simple two-threshold sequential clustering is applied to the resulting stored points and each curvature center point is assigned to a unique cluster. The clustering result can be seen in Fig. 6-6 below:

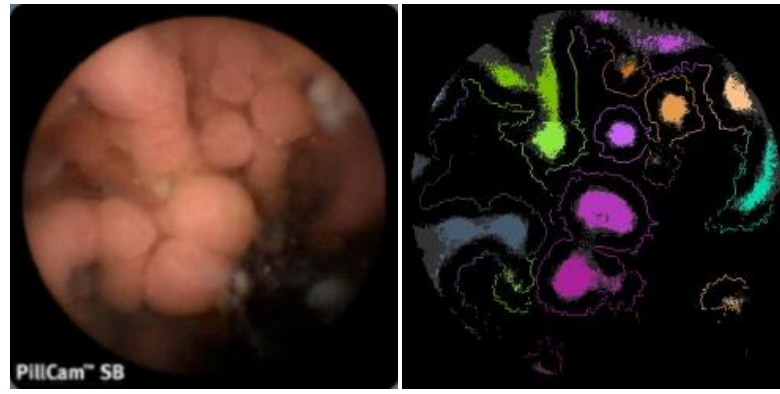


Fig. 6-6. Clusters of curvature center points with their associated boundary pixels. Each cluster has a different color

Starting from each single cluster center a level-set model (see [74]) expands an active contour on the original WCE grayscale frame until it reaches the corresponding boundary pixels of that cluster. In this way single meaningful regions are extracted.

A region is said to be a polyp candidate if it satisfies two basic rules:

A) During the contour expansion process a great percentage (>85%) of its boundary pixels have been reached.

B) Its eccentricity is above 0.70.

Rule-A makes sure that the extracted region is truly a real region with clear boundaries, while rule-B makes sure that the shape of the region is an el-

lipse close to a circle. In Fig. 6-7 six (6) cases of polyps with the results of the proposed methodology are given.

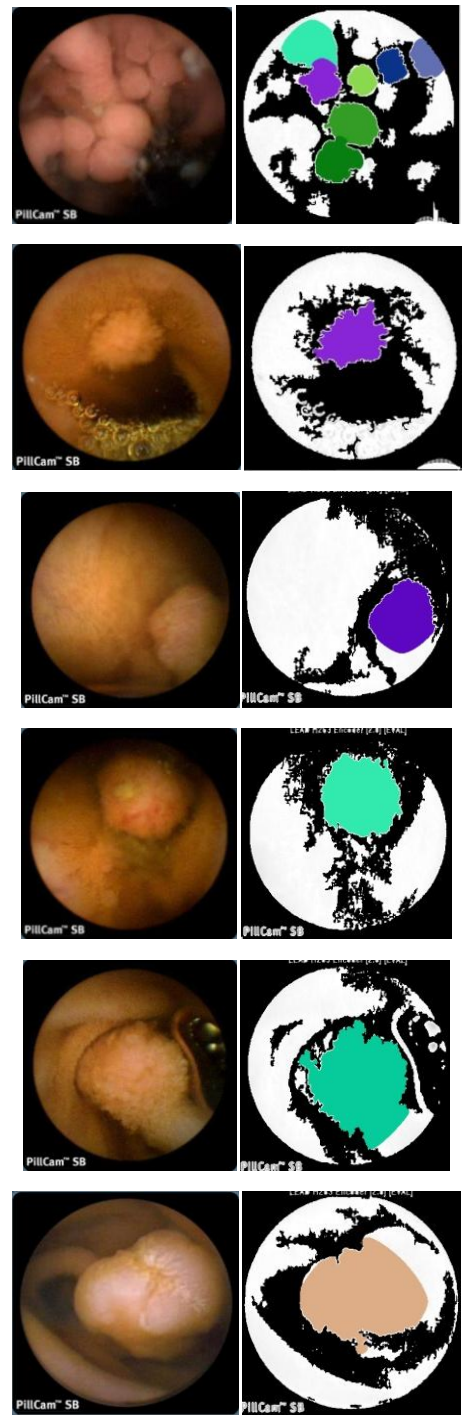


Fig. 6-7. Samples of polyps in WCE with corresponding automatic identification of them in color.

6.6 Experimental Results

The collaboration with the gastroenterologists from Digestive Associates in Dayton helped us collecting polyp cases from various patients. Some of them are shown in Fig. 6-7. As it was mentioned in section 5.3 it is really hard to gather a big number of real cases of abnormalities in WCE videos since patients do not necessarily suffer from any abnormality and if they do it, they rarely carry multiple cases.

We have run the algorithms on a 50-frame WCE video containing 10 frames with polyps and 40 normal frames in order to have a rough estimate of the sensitivity and specificity of our methodology (see Table 6-1).

	Polyp Frames	Non polyp frames
Identified	10	13
Not Identified	0	27

Table 6-1. Sensitivity=100% and Specificity=67.5%

Our methodology reached a high percentage of sensitivity because of the criteria selected in section 6.5: eccentricity and boundary pixels percentage. In order to further examine the tuning of these important parameters we ran two (2) sets of seven (7) different tests. Fig. 6-8, Fig. 6-9, Table 6-2 and Table 6-3 synopsize the sensitivity-specificity for various values of the parameters.

	Sensitivity	Specificity
Boundary pixels>0.50, E=0.10	68.34	85.34
Boundary pixels>0.50, E=0.20	71.43	82.54
Boundary pixels>0.50, E=0.30	77.43	80.12

Boundary pixels>0.50, E=0.40	82.32	77.31
Boundary pixels>0.50, E=0.50	89.32	74.43
Boundary pixels>0.50, E=0.60	94.68	72.65
Boundary pixels>0.50, E=0.70	96.23	70.23

Table 6-2. Sensitivity VS Specificity for constant boundary pixels percentage

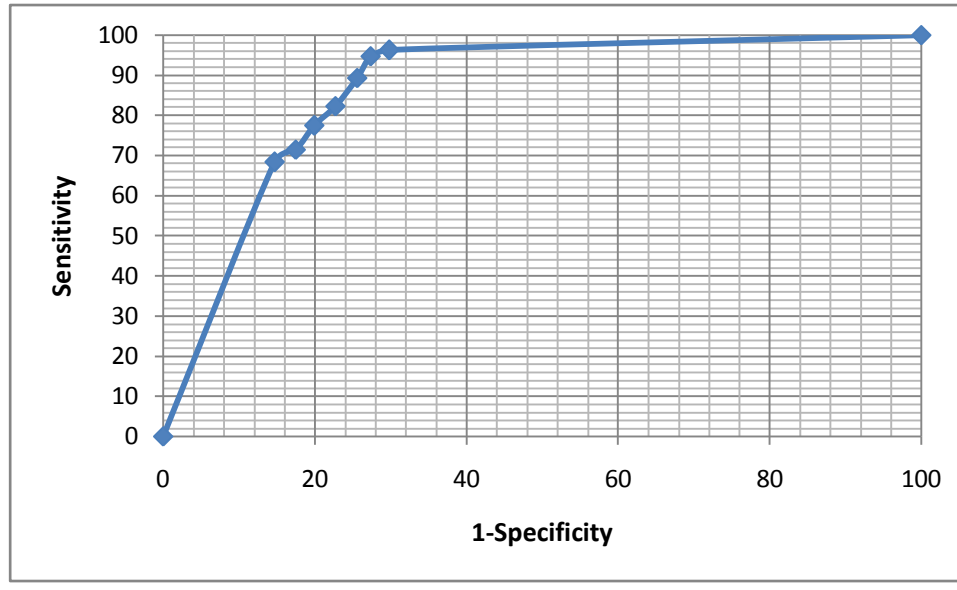


Fig. 6-8. ROC for constant boundary pixels percentage experiments

	Sensitivity	Specificity
Boundary pixels>0.55, E=0.70	96.75	72.45
Boundary pixels>0.60, E=0.70	97.76	71.23
Boundary pixels>0.65, E=0.70	98.78	70.13

Boundary pixels>0.75, E=0.70	99.23	69.23
Boundary pixels>0.80, E=0.70	99.75	67.72
Boundary pixels>0.85, E=0.70	99.96	67.52
Boundary pixels>0.90, E=0.70	100.00	67.32

Table 6-3. Sensitivity VS Specificity for constant eccentricity

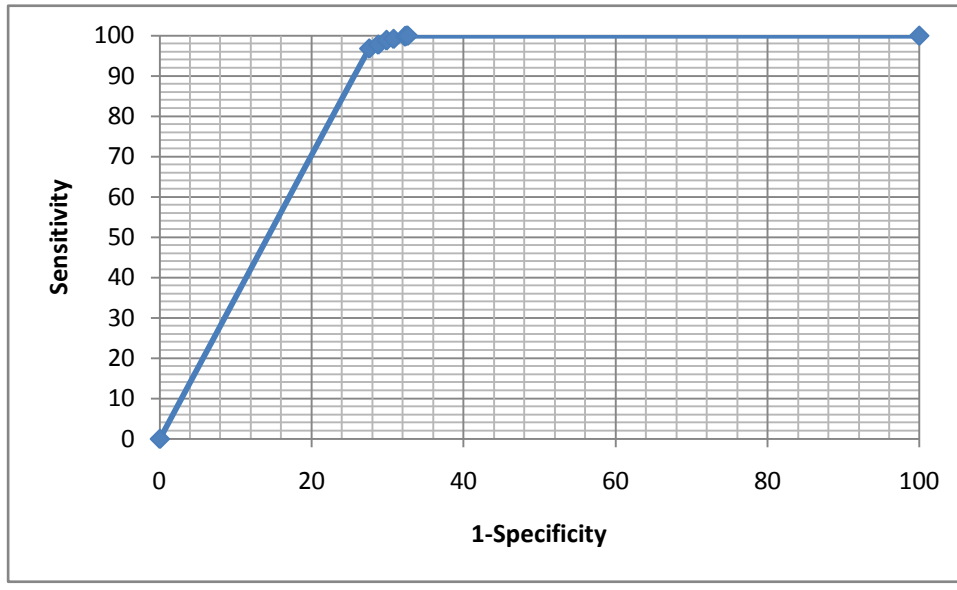


Fig. 6-9. ROC for constant eccentricity

We have performed an exhaustive search in the scientific community for works related to the detection of polyps in WCE videos because we wanted to compare our methodology with other methodologies. Unfortunately, we could not find methodologies exclusively focused on the detection of polyps in WCE. However we were able to find [75], [54] and [129] that try to identify any type of tumors. In all three (3) works researchers used texture descriptors to

describe the texture of the tumors and detection was performed using classifiers, (Table 6-4).

More specifically in [75] authors used texture unit transformation (NTU) to recognize characteristic relative positions of pixel. In such way texture units characterize the local texture information for a given pixel and its neighborhood, and the statistics of all the texture units over the image show the global texture aspects.

In [54] researchers proposed the use of Color Wavelet Covariance (CWC) Features to describe tumor texture in endoscopic images. In more detail, they perform 2-D discrete wavelet transform on each channel of the image (R, G, B) then calculate texture measures and finally these are fed to a classifier for detection.

In [129] the authors instead of using a Wavelet transform they utilize the Discrete Curvelet Transform (DCT). Having extracted the DCT coefficients they produce the textural measure that are fed in traditional MLP neural network.

	Sensitivity	Specificity
Texture units [75]	0.9714	0.9428
Color Wavelet Covariance (CWC) [54]	0.9360	0.9930
Discrete Curvelet Transform (DCT) [129]	0.9750	0.9740
Geometric characteristics (proposed method)	0.9675	0.7245

Table 6-4. Comparison of polyps detection methodologies

Our work is "custom-made" for detecting polyps, which is a subcategory of tumors, and it emphasizes on the geometric characteristics of the po-

lyps instead of texture which is differs a lot from case to case.

This approach is based on the robust and promising log Gabor filters, which behave successfully as a segmentation scheme. With log Gabor filters we managed to extract the dominating texture segments and leave out the background "less meaningful" textures, which is the goal of finding polyp candidates in WCE images. Also the SUSAN edge detector, as shown here, provides details for the boundaries, of the regions, while the geometric characteristics like curvature and eccentricity produce the final polyp candidates.

The relatively low specificity shown in Table 6-1 derives from the fact that crisp regions exist in many WCE frames, which can be easily mistaken for polyp candidates. However, it can be argued that it is up to the gastroenterologist to decide whether a polyp candidate is truly a real polyp or just a normal crisp region of the intestine. But even with this strong argument, it is our responsibility to further reduce the false positives.

Thus, to increase specificity, we are in the process of adding more sophisticated rules that engage roughness, and irregularity of the boundary. Further features like texture and color in comparison with the rest of the frame can lead to better results. Finally we are developing a fuzzy SVM to be trained with all these features in order to increase robustness and reliability.

CHAPTER 7: IDENTIFICATION OF SMALL BOWEL ULCERS

A peptic ulcer is an area where tissue has been destroyed by gastric juices. Gastric juices are produced by the stomach and the intestine to digest the starch, fat, and protein in food. Since the intestine and the stomach also consist of proteins, they are protected by a) mucous layer, b) bicarbonate, which neutralizes acid and c) prostaglandins, which are hormones to boost bicarbonate and mucus production. In case these defense mechanisms are disturbed and therefore, acid and pepsin are allowed to attack the wall of the GI tract, ulcers may result. Peptic ulcer disease affects all age groups, but is rare in children. Men have twice the risk for ulcers as women do [60].

In the United States, an estimated 25 million people will suffer an ulcer at some point [61].

Although most peptic ulcers appear in the stomach (gastric ulcers) and the duodenum (duodenal ulcers) they may also appear in the small bowel. These small intestine ulcerations can be connected with various peptic diseases, such as lymphoma, carcinoma, and Crohn's disease. In some cases no specific causes can be found.

Basically an ulcer can turn into three (3) types of complications: a) bleeding ulcer, b) perforated ulcer and c) narrowing ulcer. A bleeding ulcer occurs when the ulcer erodes one of the blood vessels. The mortality rate for bleeding peptic ulcers is about 10% [60]. Perforated ulcers appear as a hole in

the wall and they are the typical case of peptic ulcers [62]. They can lead to intense abdominal pain. Narrowing ulcers (ulcerated strictures) cause stenosis (stricture) of the intestine and can lead to severe vomiting.

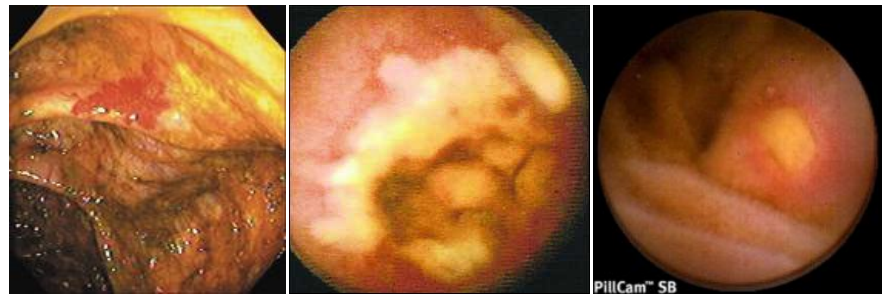


Fig. 7-1. Three (3) types of ulcers. Left: bleeding ulcer, Middle: narrowing ulcer, Right: Perforated ulcer

As it can be seen in Fig. 7-1 peptic ulcers vary morphologically. Bleeding ulcers detection can fall into the category of blood-based abnormalities ([63],[64],[65]) whereas narrowing ulcers and perforated ulcers have very different patterns so that a new detection approach should be made. To our knowledge there has been only one research work [66] exclusively on ulcers identification in WCE videos. In our research we are dealing with the detection of perforated ulcers, since these are the typical ulcers [62]. Thus, we are simply referring perforated ulcers as ulcers.

Like any image processing problem it is crucial that the preprocessing step distinguishes the target as much as possible from the rest of the image. In our case ulcers have some certain features:

a) They appear yellowish

b) They have an elliptical shape

c) They are surrounded by a redness (rush)

Feature a) is an important characteristic that can be used to extract ulcer candidates. Feature b) is an additional feature that can discard irrelevant candidates. Feature c) can be combined with feature a) to create a pattern for finding the best ulcer candidates. Texture information can be a final feature to increase the accuracy of the results.

7.2 HSV Color Space

Color is the way the human visual system measures a part of the electromagnetic spectrum, approximately between 300 and 830 nm. A color space is a notation by which we can specify colors, i.e. the human perception of the visible electromagnetic spectrum [76].

There are numerous color spaces but the following stand out: RGB, HSV, CMYK, YPbPr, Luv, Lab.

RGB color space is probably the most well known color space. It is commonly used in most of the devices for capturing images and it derives from the trichromatic theory, which states that there are three (3) types of photoreceptors, approximately sensitive to the red, green and blue region of the spectrum. RGB is an additive color space since the final color comes from the addition of its three components: red, green and blue.

Although RGB color space resembles the human visual system in kind of a similar way, it suffers from a major disadvantage: in applications with natural images there is a high correlation between its components: about 0.78 for r_{BR} (cross correlation between the B and R channel), 0.98 for r_{RG} and 0.94 for r_{GB} [76]. This was also shown by Ohta in [77] where the Karhunen-Loeve transfor-

mation of the RGB proved to achieve decorrelation of RGB components.

In 1978, Alvy Ray Smith proposed the HSL and HSV color spaces, which are very similar. HSV (Hue, Saturation, Value) color space describes colors the same way human visual system perceives colors. Hue is another word for color. It shows what color we have. Red, blue, and yellow are the primary hues, and when combined in equal amounts they create the secondary hues orange, green and violet. Saturation is the intensity of a color (or hue). Saturated colors are very pure, vivid. Value is the lightness or darkness of a color creating a scale from pure black to pure white.

The RGB-to-HSV transformation is given by:

(Eq. 7-1)
$$h = \begin{cases} 0 & \text{if } \max = \min \\ (60^\circ \times \frac{g-b}{\max - \min} + 0^\circ) \bmod 360^\circ, & \text{if } \max = r \\ 60^\circ \times \frac{b-r}{\max - \min} + 120^\circ, & \text{if } \max = g \\ 60^\circ \times \frac{r-g}{\max - \min} + 240^\circ, & \text{if } \max = b \end{cases}$$

(Eq. 7-2)
$$s = \begin{cases} 0, & \text{if } \max = 0 \\ \frac{\max - \min}{\max} = 1 - \frac{\min}{\max}, & \text{otherwise} \end{cases}$$

(Eq. 7-3) v = max

where r,g,b are the normalized values [0,1] in RGB color space.

7.3 Support Vector Machines

Support Vector Machines (SVMs) are supervised learning methods widely used to classify data. The basic concept is that an SVM maps the input data to an n-dimensional space, where it tries to find the optimal hyperplane to separate the data sets [78]. The popularity of SVMs lies on their generalization

ability for a wide range of pattern recognition problems. As well put by John Shawe-Taylor & Nello Cristianini: The key features of SVMs are i) the use of kernels, ii) the absence of local minima, iii) the sparseness of the solution and iv) the capacity control obtained by optimizing the margin.

The process of SVM is shown in Fig. 7-2 and Fig. 7-3, below.

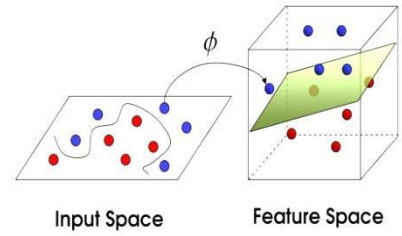


Fig. 7-2. Mapping of the input space to the feature space. Optimal hyperspace is found for the 2-class problem. Source: Council of Scientific and Industrial Research

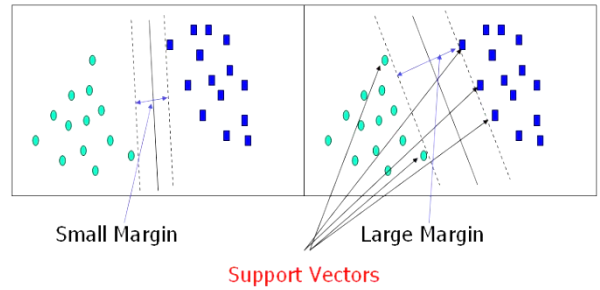


Fig. 7-3. Margin maximization using SVM. Source: DTREG.com

The optimal hyperplane that SVM finds is the one that will maximize is the margin between the support vectors. Margin is the distance between the dashed lines (see Fig. 7-3). The vectors that define this margin are called support vectors (see Fig. 7-3). Therefore, the larger the margin is the smaller the generalization error of the classifier is.

7.4 The Synergistic Methodology

In the previous paragraphs we presented each part (tool) of the methodology individually, while in this paragraph the whole methodology scheme is given. All steps of methodology are synopsized in the following diagram (Fig. 7-4).

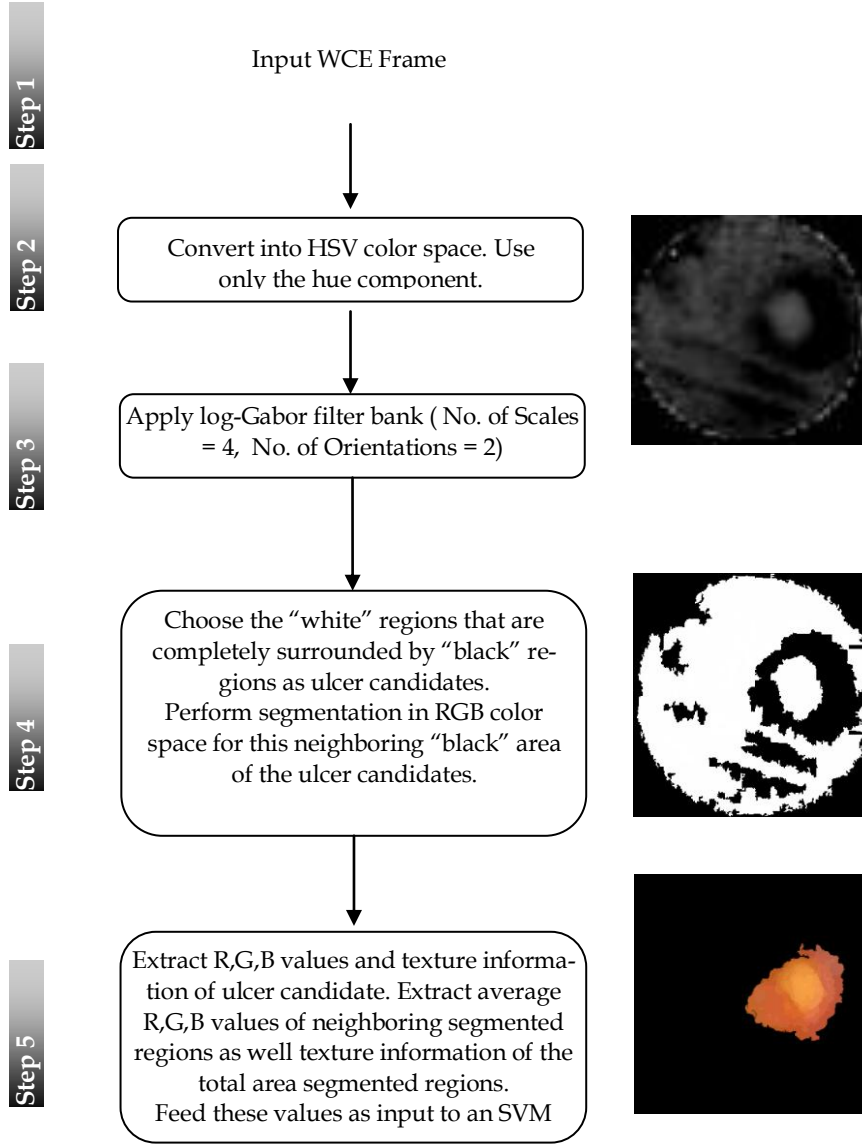


Fig. 7-4. Flowchart of the ulcer identification methodology

The Fig. 7-4 above demonstrates the whole methodology graphically. In step 2 one can easily see the importance of hue component to distinguish different colored regions in WCE frames. Hue makes regions with same color appear unified, while regions with different colors get separated. Having preprocessed the input frame with the RGB-to-HSV color transformation, log-Gabor filtering creates a robust binary segmentation of the image. One could argue whether there is a point of performing filtering with log-Gabor since a simple hard threshold on hue component could produce the same results. The answer lies exactly on that threshold. A fixed threshold value cannot cope with any given input, whereas log-Gabor filtering takes into account the whole input signal, while maximizing local frequency and spatial information.

The post log-Gabor processed frame (step 3) behaves as a forming block for the next step - the main segmentation process in RGB color space: medium-sized regions (shown in white at step 2) are ulcer candidates, so that the segmentation scheme extracts the surrounding regions, and synthesizes them [28] to produce only one big crisp surrounding region. By the term "surrounding regions" we experimentally chose regions at a fixed distance (20 pixels since the frame is 380x380 pixels) around the white area (ulcer candidate).The procedure is given below:

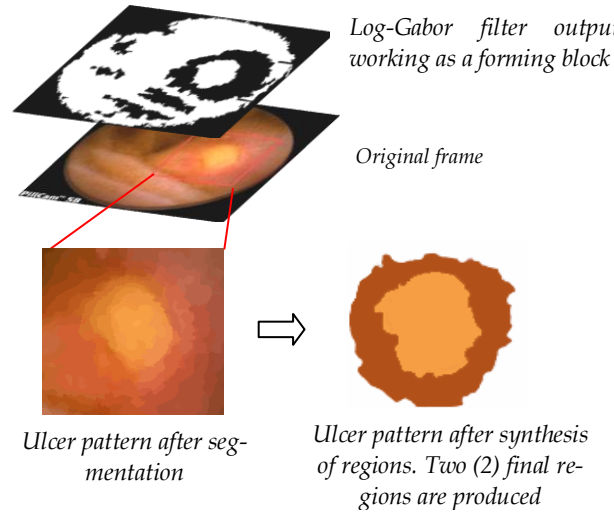
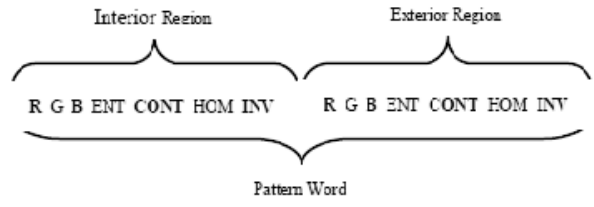


Fig. 7-5. Extraction of ulcer pattern

Having extracted a candidate ulcer pattern (inside region/hole and outside region/rush, see Fig. 7-5), average RGB values and texture information are calculated on both areas creating a "word" (description) of the ulcer pattern. Texture information consists of traditional statistical texture features described by Haralick et al. in [79]. In our case we used entropy (ENT), contrast (CONT), homogeneity (HOM) and inverse moment (INV). So we have a word format of 14 values:



The pattern word is fed to a fuzzy Support Vector Machine (SVM) which is described in [80] for classification of the input pattern.

7.5 Experimental Results

In this paragraph we present with three (3) more examples of ulcers to show how our methodology behaves with multiple ulcer candidates, Fig. 7-6. We also give some statistical results based on 20 cases of ulcers.

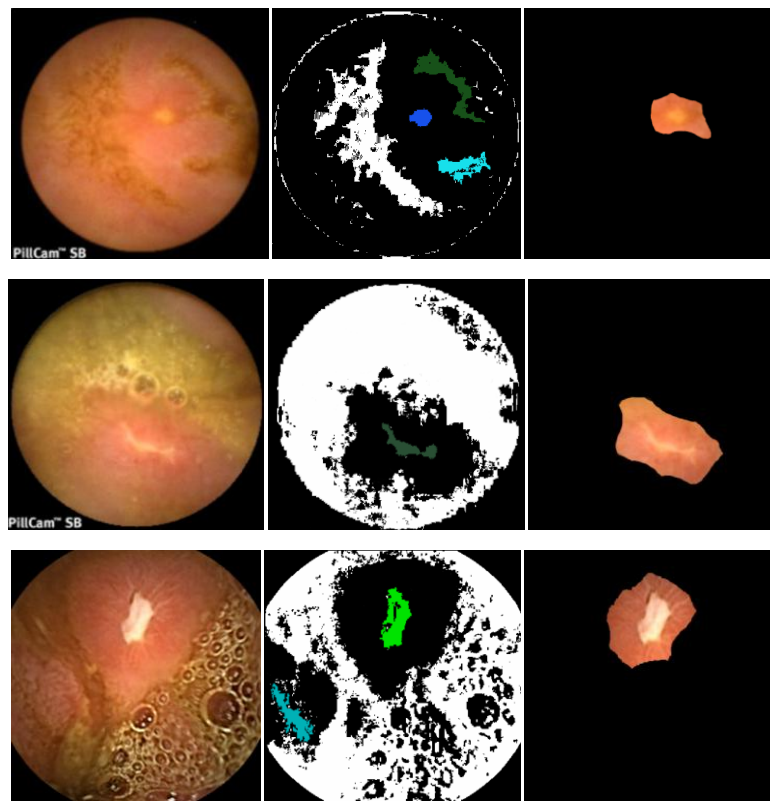


Fig. 7-6. Three cases of perforated ulcers. Left: Original Frame, Middle: Log-Gabor output (Multiple ulcer candidates are given in different colors), Right: Final output

Collecting WCE videos with cases of ulcers is quite a hard task. We managed to run our algorithm on 20 frames with ulcer cases. In addition 10 extra

frames with ulcer cases were used to train the f-SVM. We run our algorithm on those 50 frames, 20 ulcer frames and 30 non ulcer frames, in order to calculate roughly sensitivity and specificity of our methodology, Tabl2 7-1.

	Ulcer Frames	Non Ulcer Frames
Identified as ulcers	15	8
Not Identified as ulcers	5	22

Table 7-1. Sensitivity=75% and Specificity=73.3%

We searched the scientific literature for other methodologies that might deal with the detection of ulcers in WCE videos. We were able to find one work exclusive for ulcers detection [130], Table 7-2. Similarly to [128] the authors use Curvelet transform to extract the DCT coefficients. Using Local Binary Pattern (LBP) and the DCT coefficients texture descriptors are produced for the ulcers. Finally a traditional MLP neural network is used to classify LBP inputs.

	Sensitivity	Specificity
DCT transform [130]	0.8881	0.8419
Log Gabor	0.7500	0.73.30

Table 7-2. Comparison of ulcers detection of methodologies

Our methodology has focused on perforated ulcers since these are characteristic ulcers [81], whereas the authors of [130] do not clarify the type of ulcers they detect. There is a big difference in the patterns of various types of ulcers.

In the future, we plan to expand our methodology to include narrowing ulcers detection. For this direction, we used the fuzzy Support Vector Ma-

chines [80] which enables robust multiple class classification (perforated ulcers and narrowing ulcers). At the given moment, our classifier is only a two-class, either perforated ulcer or not.

We plan to gather more ulcer cases to reach a safer conclusion on the sensitivity and specificity of our methodology. This will take some time. As of now, our methodology has a good percentage of accuracy but still it needs improvement. One of the reasons that sensitivity is relatively low derives from the fact that the number of training data of the SVM was very small – only ten.

We plan to use other features for describing ulcer patterns. These features were roundness, and eccentricity. However we ran several tests and they proved not to increase accuracy. This is probably due to the irregularity of the shape of each ulcer. There is not a standard shape for all ulcers (i.e. round, elongated) and such features cannot help as they do in case of polyps. We plan to use more sophisticated texture operators such as Local Binary Pattern in order to test the performance on the accuracy.

CHAPTER 8: DATA MINING APPROACH ON POLYPS AND ULCERS

In this chapter we present WCE results by utilizing data mining techniques to extract useful information from a dataset of abnormal regions and non-abnormal regions. The goal is to find similarities and differences among polyps and ulcers by using as many features as possible to describe the extracted regions: a) healthy tissue regions, b) polyps and c) ulcers.

It is ground truth that every natural object in an image can be described by three (3) characteristics:

a) Geometry

b) Texture

c) Color

Keeping this in mind we tried to find features that offer as much information as possible for these characteristics. Concerning first category, geometry, there are many shape descriptors that can help measure regions. After careful investigation we used the following statistics: Area, Coordinates of region centroid, Major Axis Length, Minor Axis Length, Eccentricity, Orientation, Convex Hull, Equivalent diameter, Solidity, Extent, Perimeter. Since shapes in nature are not perfect circles but rather irregular major axis length, minor axis length and eccentricity describe a hypothetical ellipse that best fits the region. Convex hull or convexity is a feature that offers a degree of roughness of the region's periphery. Equivalent diameter computes the diameter of the circle

that has the same area as the region. Solidity is a measure of the firmness/ compactness of the region. Extent gives a measure of how widely spread the region is inside a boundary box.

For the second category, texture, we used some of the texture descriptors proposed by Haralick. In his work [79] he describes fourteen (14) statistics that can be calculated from the co-occurrence matrix of the image. In our case, we utilized four (4) of these measurements: entropy, contrast, homogeneity and inverse moment.

For the last category we simply used the average values of red, green and blue components of the extracted regions. Additionally we calculated the average grayscale value of the regions.

Thus, we came up with a large multidimensional space (20-feature space) to describe the regions efficiently.

8.1 Data Mining – Information Extraction

Data mining is a process that helps discover uncovering patterns, associations and structures inside data. With data mining we try to find correlations or patterns among features stored in databases.

Before we are able to perform data mining we need to create a dataset with the three (3) types of tissues. Based on our methodologies, we extracted from various WCE videos (totally 3000 frames) forty (40) ulcer regions, forty (40) polyps regions and a hundred twenty (120) healthy tissue regions that were potentially either polyp candidates or ulcer candidates. Thus, the total

measurements were four thousand (4000). A simple flowchart that demonstrates our data mining is given below in Fig. 8-1.

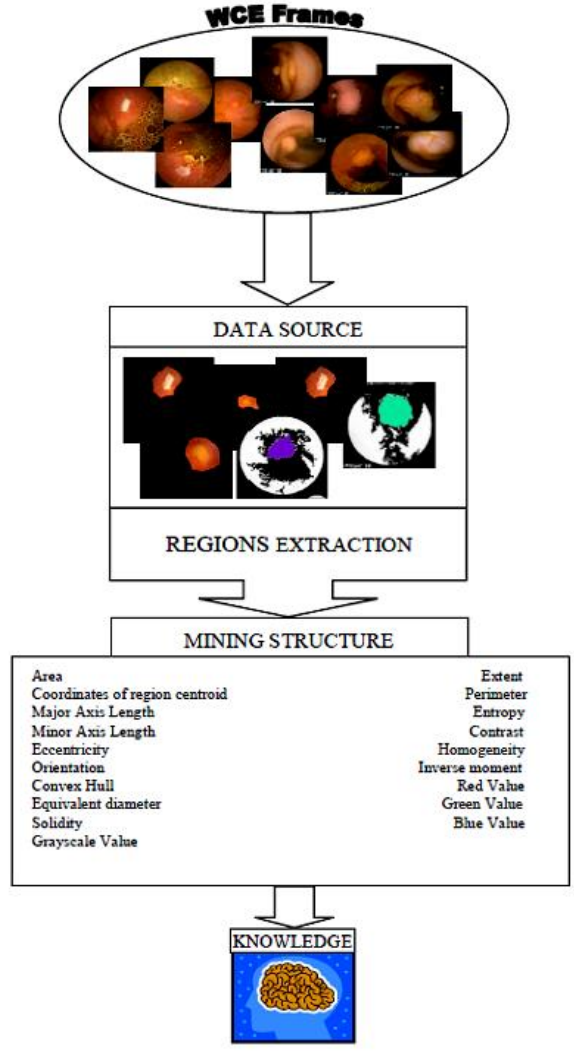


Fig. 8-1. Data mining flowchart

As data mining software we used WEKA 3-6-0 developed by the University of Waikato [82]. The WEKA toolbox offers various options: preprocessing, classification, clustering, features associating and of course visualization. Using this toolbox we were able to perform experiments on our dataset

easily and efficiently.

8.2 WCE Experiments

We have performed two (2) types of experiments with our dataset. In the first experiment we used the whole dataset to try to find correlations between the three (3) types of tissues. In the second experiment we used only the two (2) abnormal tissues. Both experiments had two phases: i) classification and ii) associations. In classification step we evaluated various classifiers to find those which could classify the tissues types more accurately. This was part of an ongoing effort to increase the accuracy of detection of polyps and ulcers. In associations step we tried to find hidden correlations between tissues.

8.2.1 Polyps vs. Ulcers vs. Normal Tissues

The WEKA software offers a wide range of classifiers to use. In table 8-1 we are presenting the top ten (10) classifiers that performed well. As one can see from that table, nearest neighbor classifier (Fig. 8-2) behaved the best for all categories of tissues. IB1 classifier worked perfectly only for ulcers whereas Simple Logistic classifier worked best for healthy tissues. It can also be seen that polyps' detection didn't reach sensitivity above 70%. The reason for this is that polyps have shape, texture and color very similar to normal tissues, making it hard to distinguish them. On the other hand, ulcers reached a great level of sensitivity and specificity which can be explained mainly due to the difference in texture and color from polyps and healthy tissues.

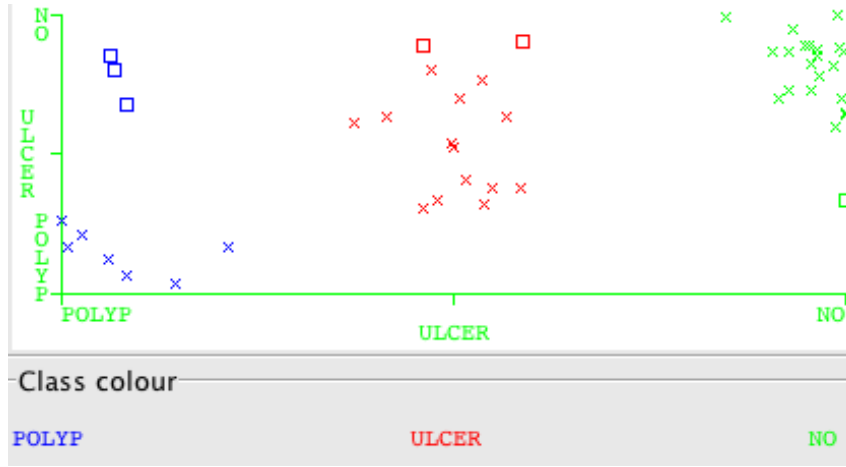


Fig. 8-2. Nearest Neighbor with generalization rule is the toppest classifier in our dataset. Xs are instances that were assigned to the correct category whereas □ are instances that were assigned to incorrect category

Polyps

	True Positive	False Positive	Precision
Nearest Neighbor with generalization rule	0.7	0.0	1.0
Logistic model tree	0.6	0.051	0.75
Naive Bayes/Decision-Tree	0.5	0.077	0.625
Fuzzy Lattice Reasoning (FLR) classifier	0.4	0.0	1
Voting feature intervals classifier	0.7	0.179	0.5
Classification via Regres-sion	0.5	0.051	0.714
Instance-based learning (IB1) classifier using 1 nearest neighbor(s) for classification	0.6	0.051	0.75
KStar Beta Verion	0.7	0.256	0.412
NN Multilayer Perceptron	0.6	0.026	0.857
Simple Logistic classifier	0.6	0.051	0.75

Ulcers

	True Positive	False Positive	Precision
Nearest Neighbor with generalization rule	0.882	0.031	0.938
Logistic model tree	0.941	0.031	0.941
Naive Bayes/Decision-Tree	0.941	0.031	0.941
Fuzzy Lattice Reasoning (FLR) classifier	0.882	0.0	1
Voting feature intervals classifier	0.941	0.031	0.941
Classification via Regression	0.824	0.0	1
Instance-based learning (IB1) classifier using 1 nearest neighbor(s) for classification	1	0.0	1
KStar Beta Verion	0.588	0	1
NN Multilayer Perceptron	1	0	1
Simple Logistic classifier	0.941	0.031	0.941

Healthy tissues

	True Positive	False Positive	Precision
Nearest Neighbor with generalization rule	0.955	0.185	0.808
Logistic model tree	0.864	0.185	0.792
Naive Bayes/Decision-Tree	0.864	0.185	0.792
Fuzzy Lattice Reasoning (FLR) classifier	1	0.296	0.733
Voting feature intervals classifier	0.636	0.148	0.778
Classification via Regression	0.909	0.296	0.714
Instance-based learning (IB1) classifier using 1 nearest neighbor(s) for classification	0.909	0.148	0.833
KStar Beta Verion	0.636	0.296	0.636
NN Multilayer Perceptron	0.955	0.148	0.84
Simple Logistic classifier	0.864	0.185	0.892

Table 8-1. Sensitivity, (1-specificity) and precision for each tissue

Furthermore, we tried to find various associations between the features of the patterns. WEKA offers the ability to view 2D graphs between two (2) features for all classes and also identify any associations.

Graphs are very helpful since they are able to display correlations that could help distinguish the classes. We knew beforehand that eccentricity plays a major role in detection of polyps but combined with contrast feature it proved to increase the sensitivity of this detection (Fig. 8-3).

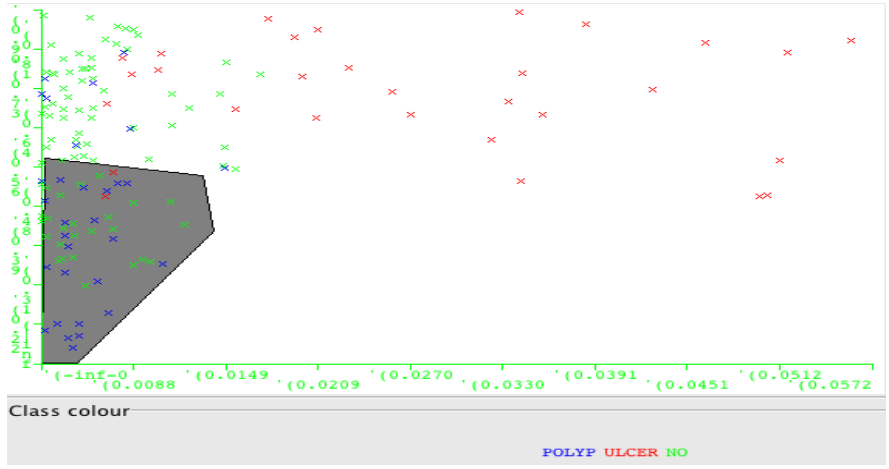


Fig. 8-3. Contrast vs. Eccentricity (shadow area shows the majority of polyps)

Elaborating on why the contrast feature helped above in Fig. 8-3, we can say that this is reasonable since contrast makes an object distinguishable from other objects (background). We moved this thinking a step further and we had a discussion with our associate physician Dr. Pouagere. From this discussion we understood that besides the shape of the polyps (eccentricity), physicians, without realizing it but due to their experience, use the concept of con-

trast to recognize polyps and extract them from the rest of the WCE frame.

Encouraged by this discovery we further investigated other features. In case of the extent feature we noticed that polyp instances are more concentrated in a certain area than the rest of the two classes (Fig. 8-4).

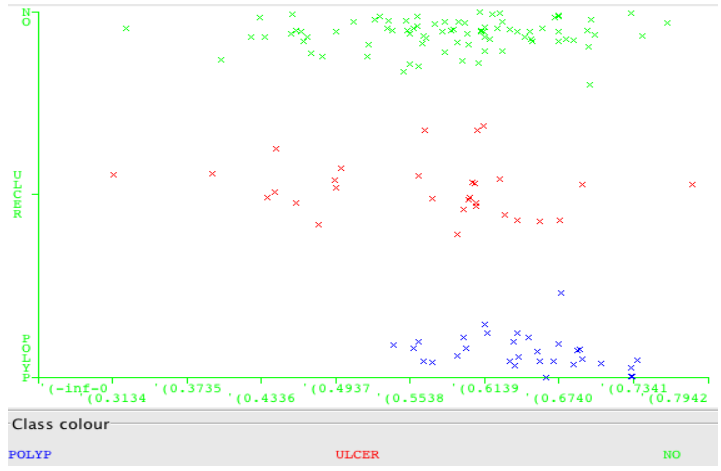


Fig. 8-4. Classes vs. Extent

This motivated us to search for other features that could help distinguish polyps. In our case extent vs. convexity increased the sensitivity of the detection. This can be explained to the fact that having applied our algorithm the extracted polyps are crisper regions with distinct borders. Since these regions are created naturally it is really difficult that they will be round shaped. On the contrary, they have protrusions and intrusions creating concave areas and thus higher degree of convexity (Fig. 8-7). Fig. 8-5 shows exactly that polyps have higher value in convexity.

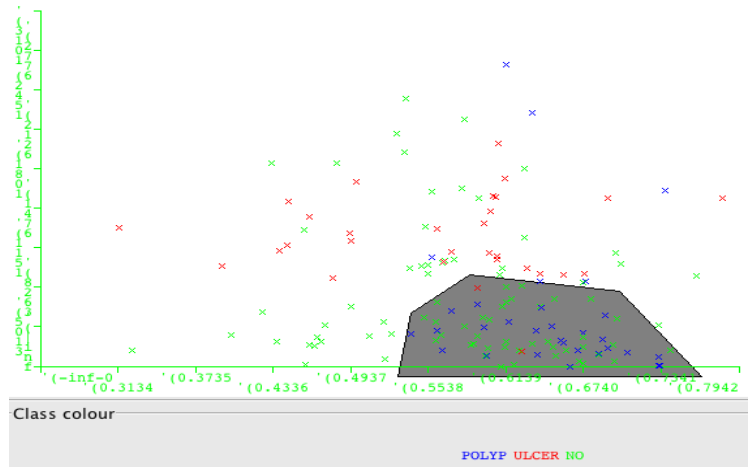


Fig. 8-5. Extent vs. Convexity

Another really useful conclusion was the coordinates of the regions of each category. Most of polyps and ulcer were concentrated near the center of the frame whereas the healthy regions were scattered all over the frame (Fig. 8-6). This shows that it is wiser for our algorithm to look inside a perimeter of the frame than scan for the whole frame for abnormalities.

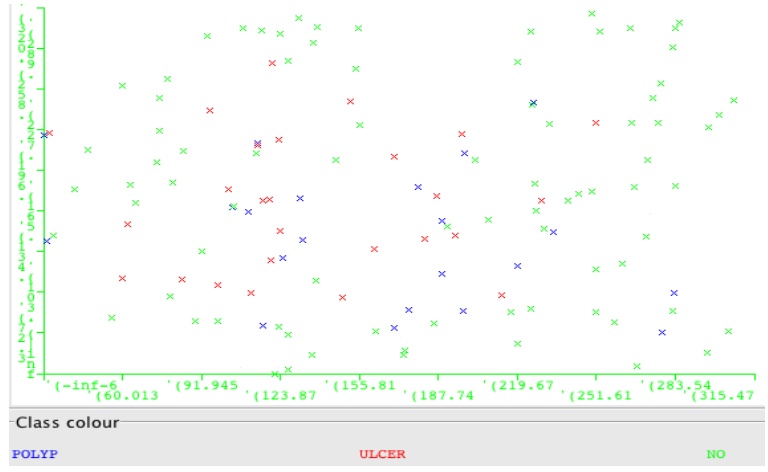


Fig. 8-6. X coordinates and Y coordinates of classes

WEKA also offers an associator tool to identify associations between

features. In our case an interesting association came up: if the instance belongs to a normal class then the inverse moment value is below 39.606 with confidence 94%. Another association is that if convexity is below 5013.9 then the area is smaller than 4571 pixels with confidence of 93%. This means that medium sized-regions have quite relatively large convexity showing the roughness of the perimeter of the region (Fig. 8-7).

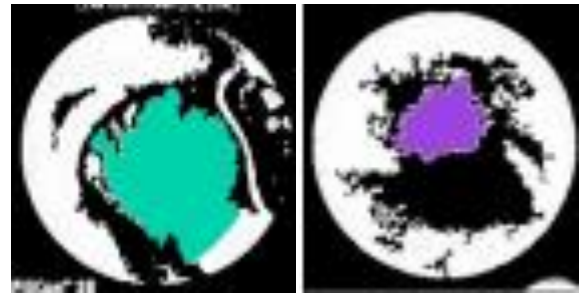


Fig. 8-7. Typical extracted polyps. Roughness of the perimeter is obvious

8.2.2 Polyps vs. Ulcers

We decided to examine the two abnormalities, polyps and ulcers, and search for any correlations and associations.

Polyps and ulcers have more differences than similarities, which is normal. Our assumption was that the polyps are more round shaped regions than ulcers. However, going through the data set we found out that ulcers were closer to circle shape than polyps (Fig. 8-8).

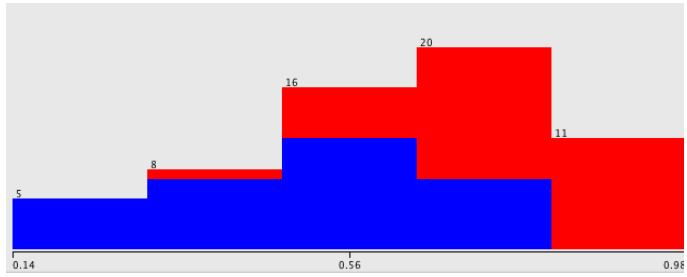


Fig. 8-8. Eccentricity feature for polyps (blue) and ulcers (red)

Surprisingly enough ulcers proved to cover averagely a bigger area than polyps (Fig. 8-9). After all, that is well understood since ulcers consist of two areas (inner and outer areas, see [83])

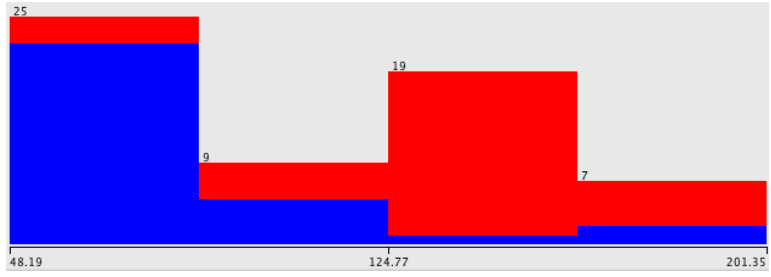


Fig. 8-9. Equivalent diameter feature for polyps (blue) and ulcers (red)

Furthermore, ulcers have higher value of entropy than polyps. This can be explained from the fact that ulcers have intense texture whereas polyps are more homogenous.

Comparing two (2) features at a time we figured out that polyps have contrast and solidity values in narrow ranges, whereas ulcers have their contrast and solidity values widely spread. This makes us believe that physicians might be able to distinguish ulcers using mainly color rather than shape and texture, whereas in case of polyps it comes down to shape and contrast. Our assumption seems to be backed by the fact that ulcers have a dominant average

color value whereas polyps do not (Fig. 8-10).

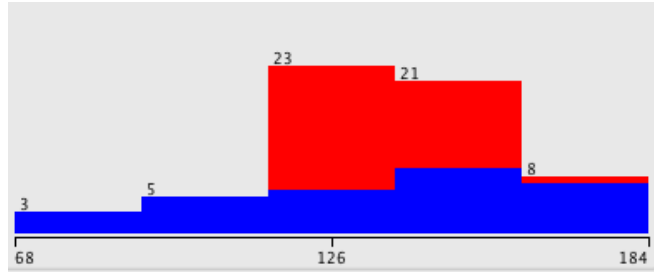


Fig. 8-10. Average color value

8.3 Conclusions

In this chapter we primarily had two goals: 1) create a big feature space and 2) evaluate these features so that to come up with meaningful conclusions about abnormal and normal regions. It is surprising to see that getting into the data sets and looking at them from every perspective led us to see some new correlations and confirm some assumptions.

Our experiments focused mainly on all three (3) classes rather than just the two (2) of them –polyps and ulcers. The reason was that it was more challenging for us to try to find ways to distinguish the three (3) classes. Besides that, the evaluation of the classifiers in 8.2.1 is going to help us implement the best classifier into our software for the detection of polyps.

Finally, our data mining helped us comprehend scientifically how physicians perceive certain patterns. Before this we lacked this knowledge and we were merely making assumptions on how medical personnel examine the WCE videos. This is probably one of the most important contributions of our data

mining: try to understand how humans perceive certain patterns and thus, try to mimic their behavior and perception to create smart algorithms.

In the future we plan to create an even larger dataset including other pathological patterns such as blood-based abnormalities. Additionally we want to expand the feature space to include time (frame number in the case of WCE videos) to investigate if there are conditional.

CHAPTER 9: 3D REPRESENTATION OF THE DIGESTIVE TRACT SURFACE IN WIRELESS CAPSULE ENDOSCOPY VIDEOS

In this chapter a proposed method is described that helps visualize the digestive tract surface in the three-dimensional (3D) space using images (2D) of WCE videos. It is not intended to reconstruct precise 3D structures of the digestive tract.

To the best of our knowledge a full real 3D reconstruction could only be feasible with the use of miniature devices/sensors that currently exist in the market or that will exist in the near future.

More specifically, the use of a magnetometer can provide not only the acceleration values on the three (3) axes but also the 3D orientation of the device (Fig. 9-1-a). Additionally, commercial time-of-flight range cameras (i.e. Microsoft's Project Natal, Fig. 9-1-b) already exist in the market and they could be improved in the near future to be implemented in an endoscopic capsule. These cameras offer not only depth information but color information too. Finally 3D guidance systems (Fig. 9-1-c) are already out in the market and used for endoscopic surgeries since they give 3D position information of the sensor.

Using the acquired information (orientation, acceleration, depth values, position, etc.) from these miniature sensors in conjunction with sophisticated registration software algorithms an accurate 3D representation of the digestive tract could be created successfully.

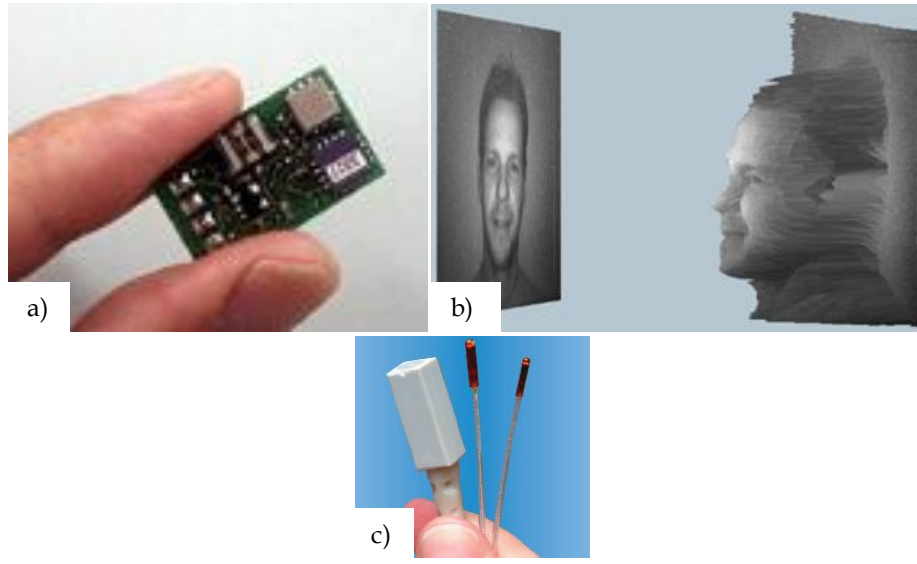


Fig. 9-1. Magnetometer, b) 3D range camera, c) 3D guidance system

9.2 Relative Work

Over the past ten (10) years there has been significant research on video endoscopy to create a 3D reconstruction of the digestive tract. Scientists have approached this challenge through hardware or software advancements.

In [84], [85], [86], stereo endoscopy systems have been proposed to capture stereo images and to create depth information and therefore 3D reconstruction of digestive structures. However due to size and compactness issues, such systems have not been widely accepted.

Software approaches are based on single images (monocular). In [87], [88], [89], [90] researchers use stereo techniques and geometric constraints from multiple frames to perform 3D reconstruction. In [91], [92], [93], [94], different structure-from-motion methodologies are proposed using consecutive video frames.

In [95], [96], [97], [98], methods that fuse different modalities (MRI, CT)

are utilized along with endoscopy to perform robust 3D reconstructions.

As far as capsule endoscopy is concerned there has been a hardware approach [99] to give in real time both 3D information and texture using an infrared projector and a CMOS camera. The major disadvantage of this system is its size and power consumption as the researchers themselves noted.

Our system falls in the category of computer aided software tools and using synergistic methods we try to construct a 3D representation of digestive tract surface by using single video capsule endoscopy frames. Thus, this chapter presents the methods used for our system, experimental results and future research directions and our conclusions.

9.3 The Methodology

A typical sequence of video capsule endoscopy frames is given in Fig. 9-2 and clearly one can see the big variety in the shape of structures. This is caused because not only camera dynamics constantly change (abrupt motion and low frame rate) but also because same regions in two consecutive frames change shape and characteristics.

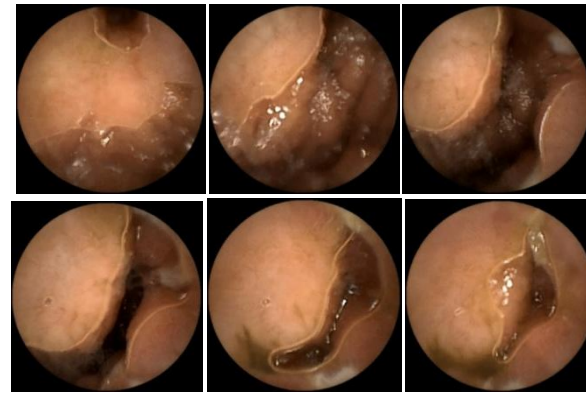


Fig. 9-2. A sequence of consecutive frames

Some motion flow techniques have been tested to create the 3D reconstruction of the digestive tract but due to the disadvantages mentioned in the previous paragraph they failed to work. For instance, the authors in [100] state that their motion flow algorithm only works with smooth transitional frames.

These facts as well as fruitful discussions with our collaborating medical doctors helped focus our research on creating a methodology that can extract the 3D structure of the surface of each frame.

9.4 Shape From Shading

The Shape from Shading (SfS) technique is a member of a family of shape recovery algorithms called shape-from-X techniques [101]. Basically shape from shading algorithm recovers the shape of objects given a single image using the gradual variation of shading (Fig. 9-3). It was first proposed by B. K. P. Horn in his PhD thesis [102].

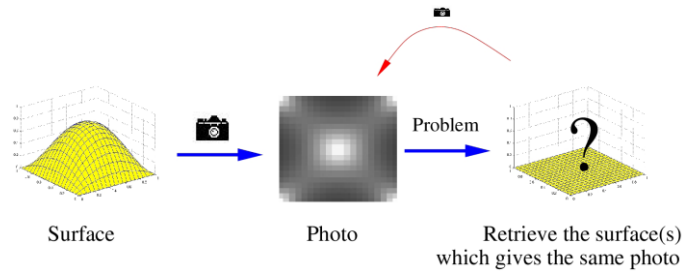


Fig. 9-3. The Shape from shading problem [103]

SfS is an ill-posed problem in the sense that additional constraints have to be applied on the surface shape parameters or the reflectance model to find the surface characteristics. The uniqueness of SFS can be proven under the condition that the light source direction is equal to, or symmetric around, the viewing direction [104].

SfS techniques can be divided into four groups: minimization approaches, propagation approaches, local approaches and linear approaches. Minimization approaches obtain the solution by minimizing an energy function. Propagation approaches propagate the shape information from a set of surface points (e.g., singular points) to the whole image. Local approaches derive shape based on the assumption of surface type. Linear approaches compute the solution based on the linearization of the reflectance map. [101]

In our methodology we use the Tsai method (linear) because it produces good results for spherical surfaces which are the case for most digestive tract shapes. It is also very fast. Additionally it behaves quite well with specular surfaces. In capsule endoscopy the light source axis and the viewer (camera) are basically aligned. Tsai method works well on smooth objects with the light

source close to the viewing direction [101] and this is because their method does not suffer from any frequency doubling effect [105].

In Fig. 9-4 the result of Tsai method is given for a single capsule endoscopy video frame.

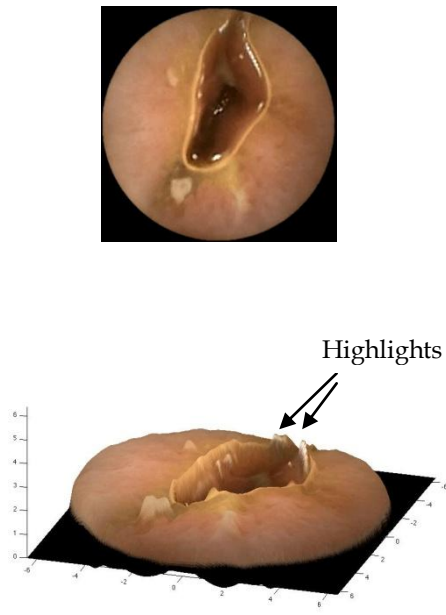


Fig. 9-4. 3D reconstruction of the surface of WCE video frame

The result is really impressive with the shape of structures (surfaces) given in detail. As expected the main downside of the resulting image is the presence of highlights.

9.5 Extracting the diffuse component

It is known that highlights are linear combinations of specular and diffuse reflection components. Many objects in the real world are dielectric inhomogeneous which have both types of reflections. In endoscopy most digestive

structures fall in this category. In case of such an object, when the light beams fall on it, some of them reflect back immediately creating the specular reflection while the rest of the beams first penetrate the object and then reflect creating the diffuse reflection.

We know that Tsai's method, as well as most SfS methods, work better on images following the lambertian reflection model. However, the intensity distribution of diffuse reflections approximately follows Lambert's law [106]. Since the Tsai method gave us really satisfying results we considered improving it by adding a reflectance model transformation as preprocessing step to make endoscopy images specular free and therefore follow the lambertian reflection model.

We used the method by Tan and Ikeuchi [107] to separate the two reflection components; diffuse and specular. This method behaves optimally since it does not require color segmentation as other methods do. Additionally another advantage is that it runs on two neighboring pixels minimizing the error of the outcome. Basically the methodology creates a normalized image containing only the pure specular component of the original image. It also creates a specular-free image out of the normalized picture by using the maximum chromaticity. A logarithmic differentiation of the two images gives the diffuse image which is the original one with the highlights removed. An overview of the methodology as well as the result after applying the algorithm on WCE frame is given in Fig. 9-5.

.

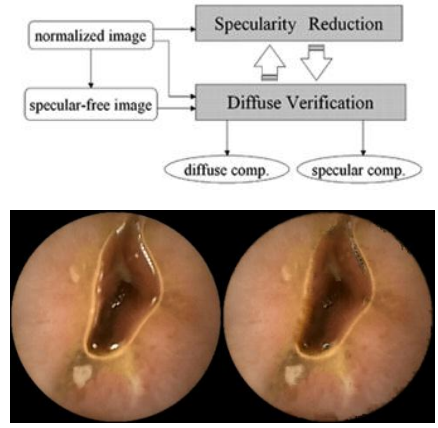


Fig. 9-5. Tan & Ikeuchi highlights removal algorithm applied on WCE image

9.6 The Registration

As explained in the beginning of this chapter since current hardware specifications do not allow us to create a 3D map of the digestive tract we had to come up with a way to align consecutive WCE frames which is the best we can do at this low frame rate. Our registration scheme, presented in detail in chapter 4, helped us to achieve this goal here.

9.7 Experimental Results

In this paragraph we are presenting results from our methodology which was developed in JavaFX ©.

The software is a viewer of the 3D representation of the WCE frames. It allows the user to load a series of the dataset and rotate each frame in 3D view using the cursor (Fig. 9-6).

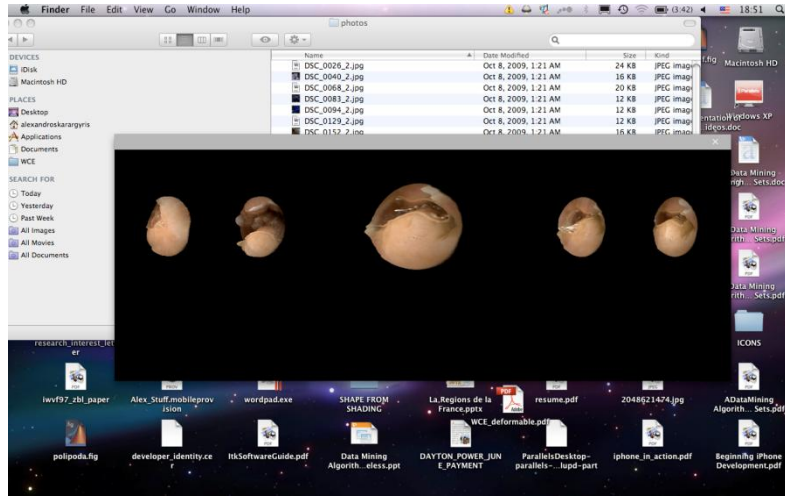


Fig. 9-6. 3D Viewer

9.8 Conclusions

In this chapter we proposed a methodology for the 3D reconstruction of the surface of WCE video frames. The methodology used an improved approach of the shape from shading algorithm to achieve far better results by removing specularities (highlights) from the images. We demonstrated the software to our collaborating physicians and they find its capabilities important. As they noted the software gave them the ability to view the patients' videos from a new perspective and 3D representation is something they have desired for a long time.

We are also working on expanding the 3D representation for a more accurate detection of polyps. We are also trying to expand our registration scheme in 3D space by using the outcome of our current research.

As discussed in the introduction an accurate 3D representation of the

surface of digestive structures as well as 3D map creation of the digestive tract can be achieved through hardware improvements. Therefore we are in the process of implementing some of the techniques briefly discussed in the introduction towards that direction. The challenges are big since the technology is not mature yet but we are confident we are on the right track.

CHAPTER 10: AN ELASTIC VIDEO INTERPOLATION METHODOLOGY FOR WIRELESS CAPSULE ENDOSCOPY VIDEOS

It is known that for a given object (i.e. ball of food, vitamin pill, unknown solid object, capsule) gastrointestinal muscles produce smooth contractions that push this object further down the tube. These smooth contractions are called peristalsis and in a given time frame they can be viewed as waves that move the object forward. However these waves are not necessarily unidirectional but bidirectional that push back and forth the object multiple times and eventually forward. So peristalsis can be seen as a forward moving vibration through the digestive tract. Fig. 10-1 demonstrates the peristalsis process in detail.

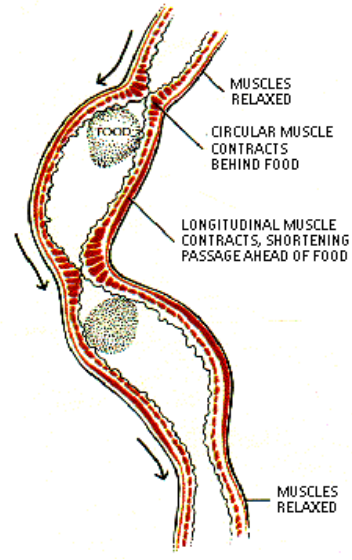


Fig. 10-1. As food traverses in the tract the muscles of the bowel contract in a wavelike fashion (relaxation and contraction) behind and ahead of the food ball

Because of this phenomenon the wireless capsule is able to move through the digestive tract. However, as explained above, peristalsis forces the capsule to bounce backwards and forwards. This causes the video inherit a complex motion instead of a smoothly forward motion. The free-form motion of capsule further complicates the already complex motion. Finally the low frame rate certainly does not improve the situation. (a WCE video sample can be found at http://www.cs.wright.edu/atrc/)

As mentioned in the introduction, the goal of this chapter is to propose a methodology that creates intermediate (interpolated) frames from two WCE frames given that these frames carry mutual information of some degree.

Our contribution consists of an advanced framework which takes into account a) the elasticity of a graph describing the structure of observed objects, b) objects deformation, c) illumination variance between the frames, d) large displacement optical flow and e) occlusions.

We will break the problem of interpolation into simple understandable steps that deal with this complex motion. All these steps, mathematically formulated, will lead to the proposed solution: creation of smooth consistent interpolated frames.

The following topics are covered: an overview of prior works, muscles contractions as an elasticity movement, non-rigid motion of completely mutual frames (with/without objects deformation – with/without illumination change), and non-rigid motion of partly mutual frames (with objects deformation – with/without illumination variance). Finally, by the term "mutual" we

mean the similarity of two frames. For reasons of ambiguity of the term similarity we preferred to use the term mutual and partially mutual information regarding the dependency of two frames.

10.2 Prior Work

Understanding the motion of objects in a sequence of images is an important problem in computer vision. Motion estimation can help in many cases, such as video interpolation. Pioneers Horn and Schunck [108] assumed that the intensity of image structures is approximately constant for at least small duration of time [109] leading to the optical flow formulation:

$$I(x,y,t) = I(x + \delta x, y + \delta y, t + \delta t) \qquad \text{(Eq. 10-1)}$$

Since then, a sufficient amount of research work has been carried out to deal with the major shortcoming of optical flow: small displacement consideration. Researchers [110]-[115] have added gradient constancy constraints and they have used various regulizers to deal with over-smoothing and edge preservation. Additional regulizers have been employed to deal with outliers. For large displacement motions warping schemes (coarse-to-fine techniques) have been utilized while recently [116] warping is avoided using an energy functional that can be globally minimized.

Although tremendous efforts have been made to optimize the energy minimization locally little work has been done to take into account the global characteristics of the image. In [117] the authors proposed a region based

matching optical flow variation. In the this chapter, we raise the importance of the global structure of the objects that consist the image using elasticity features while moving the elasticity down to the objects' parts – the pixels. Although the proposed motion flow model is proposed for the complex case of WCE videos this does not exclude it from being applied in other cases.

Finally this work also takes into account the illumination variance which is true in case of light variations in consecutive images.

10.3 Muscles Contraction-Relaxation, Elasticity and the Global Graph

As mentioned in section 10.1, the muscles contract and relax pushing the capsule forward and backward with a relative slow forward movement. Since the muscles are found behind the intestinal wall, the deformation of the wall is almost always hard to capture on the video. What may be captured on the video though is the result of this deformation which is the deformation of a graph G describing the structure of the viewed objects. To explain this let us take as an example the following consecutive segmented frames (Fig. 10-2).

Fig. 10-2. Segmented frames with visually attached graph showing the movement of the same objects (red, purple, yellow) through time. Last scheme shows this deformation on the overlapping graphs

Each graph (G_i) consists of nodes (N_i) and edges (E_i), $G_i\{\mathbf{N},\mathbf{E}\}$. Nodes are assigned to the centres of segmented regions and they carry colour, texture and

geometric properties. Edges connect the nodes and describe their relationships (distance, neighbourhood). In such a way an image can be described with a graph that gives the structure of the objects in an image. Depending on the level of detail of the segmentation process and the number of nodes we are able to tune the detail of the graph. Without loss of generality we can also define G_i at the centre of the image and the edges are adjusted accordingly (Fig. 10-2).

Another important feature of graph is that it locks a) the objects pixels to the nodes and b) the objects in a structure relationship through time.

For a given pair of frames (G_1 – G_2) let us denote the mutual graph as:

$G_m = G_1 \cap G_2$

The reader is referred to 3.3 for understanding the above intersection process. G_m carries the mutual information of G_1 and G_2.

The deformation of the muscles can be viewed as a single isotropic linear elastic body Ω undergoing a deformation. This is true since the intestine relatively consists of the same material. G_m is attached to this body and it is able to capture the result of this elastic deformation. We deliberately mentioned the "result of elastic deformation" because G_m cannot explain the physical model of the forces applied to the intestinal muscles. This however does not affect the solution to our problem. Besides, this is beyond the scope of the present chapter.

Let us define the energy of G_m as the elastic behaviour of the body Ω to be deformed:

(Eq. 10-2) $$E_{G_m} = \int_{\Omega} \sigma^t \varepsilon \, d\Omega$$

where $\sigma^t = D\varepsilon$ is the Cauchy stress vector for linear elasticity and D is the elastic matrix describing the material properties of the intestine. ε is the strain vector given by $\varepsilon = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}, \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}, \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z}, \frac{\partial u}{\partial x} + \frac{\partial u}{\partial z}\right)^t = Lu$, u is the displacement vector of the object's structure (in our case the nodal displacement U_n of G_m). For an isotropic linear elastic material the elasticity tensor D is given by the Young's elasticity modulus $\boldsymbol{E}$ and Poisson's ratio $\boldsymbol{\nu}$.

By assigning elastic properties to the graphs we firstly lock the objects of the frame into a structure and therefore they are easier to manage. Secondly, but more importantly, the elastic deformation of the graph works as an overall guide for finding a smooth interpolation. The importance of this feature will be presented later.

In the following sections different scenarios of image deformations will be presented. In each case the fitting of a right physical model and mathematical terms that best deal with each deformation will be discussed in detail.

10.4 Non-Rigid Motion With Mutual Information

We start our discussion with the simplest and most ideal motion which is the case when two frames F_t with graph G_t and F_{t-1} with graph G_{t-1} contain the same (mutual) information. By the term information we refer not only to the intensity of the pixels but also to their graphs. In this case we have:

(Eq. 10-3) $$G_m = G_t \cap G_{t-1} = G_t = G_{t-1}$$

In this case the error we want to minimize is given by:

(Eq. 10-4) $$E_{GRAPH\_DEFORMATION} = E_{G_m} + E_{ext}$$

where E_{ext} gives the applied forces that deform G_t to match with G_{t-1} and it is given by:

(Eq. 10-5) $$E_{external} = \int_{\Omega} ((G_t + V(n)) - G_{t-1})^2 d\Omega$$

With $V(n)$ describing the displacement vector of the graph nodes N_i, that need to be calculated. Substituting (Eq. 10-2) and (Eq. 10-5) in (Eq. 10-4) we have the total energy to minimize:

(Eq. 10-6) $$E_{GRAPH\_DEFORMATION} = \int_{\Omega} \sigma^t \varepsilon \, d\Omega + \int_{\Omega} ((G_t + V(n)) - G_{t-1})^2 d\Omega$$

Following optical flow concept for small deformation, therefore small displacements, this energy can be further written as:

(Eq. 10-7) $$E_{GRAPH\_DEFORMATION} = \int_{\Omega} \sigma^t \varepsilon \, d\Omega + \int_{\Omega} ((G_t + V(n)) - G_{t-1})^2 d\Omega + \int_{\Omega} |\nabla V(n)|^2 d\Omega$$

Note that the newly added term $E_{SMOOTH} = \int_{\Omega} |\nabla V(n)|^2 d\Omega$ guarantees a relative smooth flow field.

Since in WCE videos we have a low frame rate and therefore large displacement we need to follow a coarse-to-fine approach to solve the minimization problem. As with other cases we used a Gaussian pyramidal analysis in which the graph was refined at each level.

Note that in all the above equations (Eq. 10-2) - (Eq. 10-7) the integral of the elastic energy of body has been considered. That is true when the number of nodes of the graph is really big ($Ni \rightarrow AxB$, A and B the dimensions of the image). Therefore the equations need to be revised with the summation over the graph nodes ($\sum_{i=1}^{Ni}$) that consists the body Ω and (Eq. 10-7) is finally given by:

(Eq. 10-8) $$E_{GRAPH\_DEFORMATION} = \sum_{i=1}^{Ni}[\sigma^t \varepsilon + ((G_t + V(n)) - G_{t-1})^2 + |\nabla V(n)|^2]$$

10.5 Non-Rigid Motion With Deformable Objects And Mutual Information

In the previous sections we introduced the concepts of graph elasticity and its deformation capability. However we have not taken into account the deformation of each individual object (described by a graph node N_i). These local deformations can be caused by physical applied forces (muscles contraction) or camera distortions. We update the definition of graph G given in section IV so that each node N_i can carry additional information about elastic properties. In our case the elastic properties of each node are the same and this agrees with our assumption that intestinal parts consist of same material.

The energy minimization can now be written as:

(Eq. 10-9) $$E_{MUTUAL} = a * E_{GRAPH\_DEFORMATION} + (1 - a) * \sum_{i=1}^{Ni} E_{OBJECT\_DEFORMATION}$$

Now the energy is distributed among the energy needed to deform the graph and the energy needed to deform each object and this can be tuned with the parameter **α**. (Eq. 10-9) can be written in detail as:

(Eq. 10-10) $$E_{MUTUAL} = a * \sum_{i=1}^{Ni}\left[\sigma_{Graph}^{t}\varepsilon_{Graph} + ((G_t + \boldsymbol{V}(n)) - G_{t-1})^2 + |\boldsymbol{\nabla V}(n)|^2\right] + (1 - \mathrm{a}) * \sum_{\mathrm{i}=1}^{\mathrm{Ni}}\left[\int_{\Omega i}\sigma_{node}^{t}\varepsilon_{node}d\Omega_i + \int_{\Omega i}(\mathrm{F_t}(\mathrm{x} + \boldsymbol{V}(n) - \boldsymbol{U}(\mathrm{x})) - \mathrm{F_{t-1}}(\mathrm{x}))^2 d\Omega_i + \int_{\Omega i}|\nabla \mathrm{F_t}(\mathrm{x} + \boldsymbol{V}(n) - \boldsymbol{U}(\mathrm{x})) - \nabla \mathrm{F_{t-1}}(\mathrm{x})|^2 d\Omega_i\right]$$

with:

σ_{node}^{t} : nodal stress vector

ε_{node} : nodal strain vector

F_{t-1} , F_t : intensity values of frames t-1 and t respectively

$U(x)$: is the local displacement inside the deforming node (object) at point x

$\nabla(\cdot)$: the gradient $\frac{\partial}{\partial x}$

Vector $\boldsymbol{U}(\mathrm{x})$ describes the local displacement of pixels inside the objects (nodes) whereas the $\boldsymbol{V}(n)$ still describes the nodal displacements which are usually larger and affect all the pixels of each object.

$\boldsymbol{U}(\mathrm{x})$ can be solved using various point methods such Finite Element Method (FEM), Boundary Element Method (BEM) , B-splines.

The most important feature of (Eq. 10-10) is that it shows how energy travels from the graph ($E_{GRAPH\_DEFORMATION}$) to its nodes ($\int_{\Omega i}\sigma_{node}^{t}\varepsilon_{node}d\Omega_i$) and

then to the pixels of the image (last three terms: $\int_{\Omega i}(\mathrm{F_t}(\mathrm{x}+\boldsymbol{V}(n)-\boldsymbol{U}(\mathrm{x}))-\mathrm{F_{t-1}}(\mathrm{x}))^2 d\Omega_i+\int_{\Omega i}|\nabla \mathrm{F_t}(\mathrm{x}+\boldsymbol{V}(n)-\boldsymbol{U}(\mathrm{x}))-\nabla \mathrm{F_{t-1}}(\mathrm{x})|^2 d\Omega_i$). Therefore this equation manages to connect the top information with the lowest information.

10.6 Non-Rigid Motion With Deformable Objects And Partial Mutual Information

So far we have discussed the case when two consecutive frames F_{t-1}, F_t share exactly the same mutual information. In most real cases this is not true. It is better for us to redefine the information that a frame F_t carries in coordinance with another frame as:

$$I_{TOTAL} = I_{MUTUAL} + I_{NON\_MUTUAL} \quad \text{(Eq. 10-11)}$$

In the previous sections we dealt with the first term and we tried to minimize $\boldsymbol{F_{t-1}} \rightarrow \boldsymbol{F_t}$ based on this term. The second term of (Eq. 10-11) describes the information that exists in F_t and not necessarily in F_{t-1} and vice versa. We avoided using the term *occluded information* as other authors use because occlusion information refers to totally inexistent information. In our case non mutual information consists with occluded pixels but also with pixels that are not part of the graph. The above concepts are presented in Fig. 10-3.

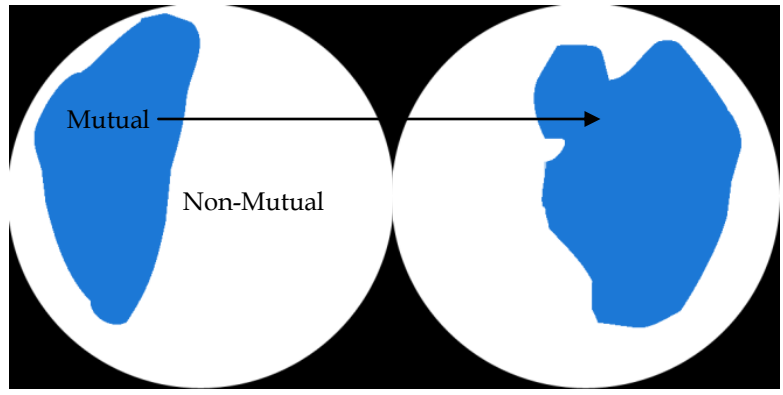


Fig. 10-3. Mutual and non-mutual information of a frame F_{t-1} in coordinance with frame F_t. Note that the blue region (mutual information) can be interpolated using the graphs

We have successfully broken the problem of video interpolation into to two parts: a) video interpolation for mutual information and b) video interpolation for non-mutual information. Mutual regions and non-mutual regions share one important component: *consistency*. This basic notion however will help us deal with non-mutual regions since mutual regions will be our guide for estimating the motion flow.

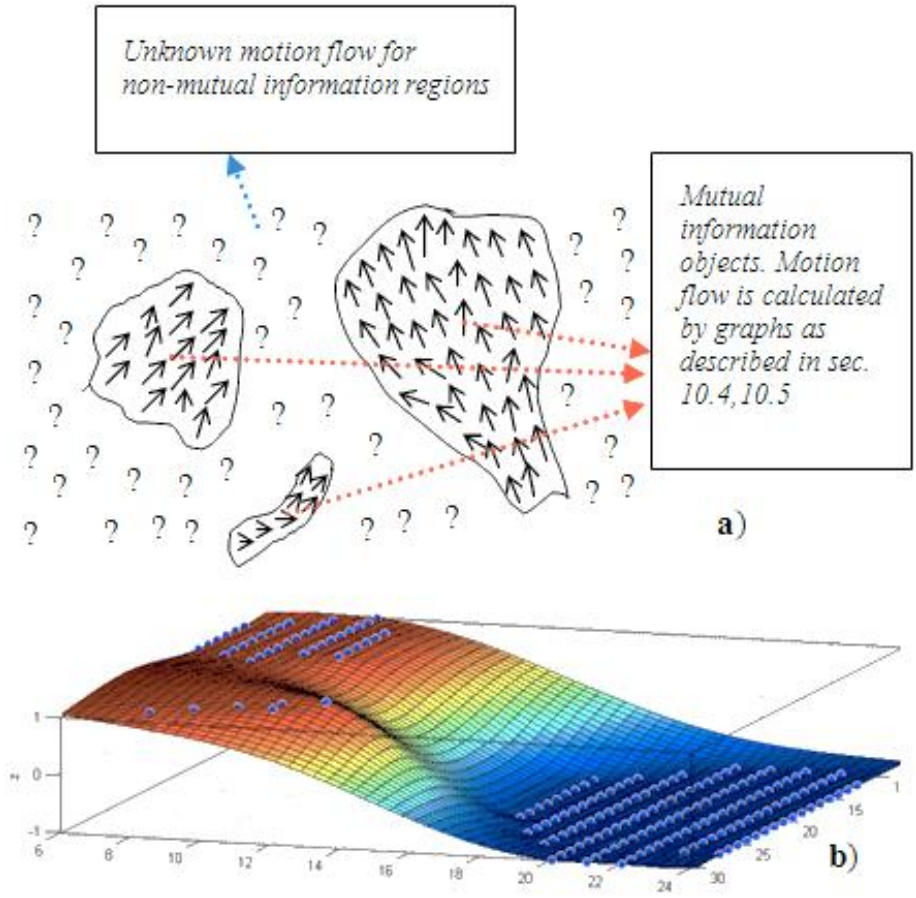


Fig. 10-4. a) An example of mutual and non-mutual regions. Motion flow is known for mutual regions through graph deformation and unknown for non-mutual regions. b) Estimation of the motion flow over the whole image plane using bi-harmonic interpolation for the Uy vectors.

With distributing the motion flow (Fig. 10-4) across the image plane we are able to suppress motion flow uncertainties in areas of sparse textured areas. Note again that this dense interpolated motion flow works as a guide and not as the ground truth, which would lead to inaccurate results, as it is discussed below.

Synopsizing all the above notions the energy we wish to minimize is given by:

(Eq. 10-12) $$E_{NON\_MUTUAL} = \int_{\Omega'} (\mathrm{F_t}(\mathrm{x} - \boldsymbol{U}(\mathrm{x})) - \mathrm{F_{t-1}}(\mathrm{x}))^2 d\Omega' + \int_{\Omega'} \left| \nabla \mathrm{F_t}\left(\mathrm{x} - \boldsymbol{U}(\mathrm{x})\right) - \nabla \mathrm{F_{t-1}}(\mathrm{x}) \right|^2 d\Omega' - \int_{\Omega'} \frac{<U(x), V(x)_{INTERPOLATED}>}{\|U(x)\| \|V(x)_{INTERPOLATED}\|} d\Omega'$$

Where $V(x)_{INTERPOLATED}$ is given by E_{MUTUAL} through interpolation. Note that the first two terms in (Eq. 10-12) are the optical flow constraints. The last term denotes the cosine of the angle between $U(x)$ and $V(x)_{INTERPOLATED}$ which we want to maximize (or minimize the negative) and it favors displacements that follow the dense motion flow allowing a smooth consistency with the mutual regions. $<,>$ denotes the inner product of vectors and $\|...\|$ the length of vector. Ω' refers to the non-mutual regions.

10.7 Illumination Variant Interpolation

So far we have discussed the motion of two frames without taking into account the variance of illumination. This is important since the aforementioned models do not consider changes in the intensity of the pixels which occur in most real cases as well as video endoscopy. The classical optical flow approach directs that I(x) = I(x+u) where x is the pixel coordinates and u the displacement vector. However in most cases I(x) ≠ I(x+u).

Basri and Jacobs [118] proved that an image can be decomposed by an infinite series of spherical harmonics functions. Under any lighting conditions, at least 98% of the variability in the reflectance function of the image is captured by the first 9 harmonic images. Therefore images of a convex Lambertian surface will

lie close to a 9D subspace.

In this case the intensity of a pixel P is given by:

$$(\text{Eq. 10-13}) \qquad I(P) = \sum_{i=0,1,2} \sum_{j=-i,-i+1,\dots i-1,i} l_{ij} b_{ij}(n_P)$$

where l_{ij} is the illumination coefficient determined by the illumination direction , and $b_{ij}(n_P)$are the basis images given by:

$$(\text{Eq. 10-14}) \qquad bij(n) = \rho r_i Y_{ij}(n)\,, i = 0,1,2; j = -i, \dots i$$

where ρ is the albedo at pixel P, r_i is constant for each spherical harmonics order, Y_{ij} is the spherical harmonics function, and n is the normal at pixel P. In such way for a given object with fixed albedo ρ we can show that its intensity at x depends only on: a) the light direction (angle between light beam and P) and b) the spherical harmonics of the normal of P. Therefore, using tensor notation, we can write for a rigid motion **U**:

$$\mathrm{I}_1(\mathbf{X}) = \mathbf{lb}$$

$$(\text{Eq. 10-15}) \qquad \mathrm{I}_2(\mathbf{X} + \mathbf{U}) = (\mathbf{l} + \Delta\mathbf{l})\mathbf{b}$$

Basis images **b** have not changed since ρ and n are constant. Therefore for a fixed light intensity the change of image intensity I is caused by a change in illumination due to the motion **U** (change of direction). Therefore $\Delta \boldsymbol{l} = f(\boldsymbol{u})$.

We are going to revise the graph G so that each node Ni can carry the

following information for its pixels: a) illumination coefficients l_{ij} and b) basis images$b_{ij}(n_P)$. Note that calculation of $b_{ij}(n_P)$ requires knowledge of the albedo and the surface normal. These are calculated using Shape-of-Shading techniques. Reader is referred to [119] for further reading. Thus, l_{ij} is calculated by unfolding basis images B and image I [120],[121]:

$$\text{(Eq. 10-16)} \qquad \hat{\boldsymbol{l}} = (\boldsymbol{B}_{(1)}\boldsymbol{B}_{(1)}^{T})^{-1} \times_1 I_{(1)}^{T}$$

Now in case of a non-rigid motion of the graph the total energy to be minimized can be given as:

$$\text{(Eq. 10-17)} \qquad \mathrm{E_{GRAPH\_DEFORMATION}} = \int_{\Omega} \sigma^{t}\varepsilon \, d\Omega + \int_{\Omega} ((\mathrm{G_t} + \mathbf{V}(\mathrm{n})) - \mathrm{G_{t-1}})^2 d\Omega + \int_{\Omega} (\mathrm{Gl}(\mathrm{n} + \mathbf{V}(\mathrm{n})) - \mathrm{Gl}(\mathrm{n}))^2 d\Omega + \int_{\Omega} |\boldsymbol{\nabla}\mathbf{V}(\mathrm{n})|^2 d\Omega$$

where Gl() returns the illumination coefficients $\boldsymbol{l}$ for node n.

We added the following energy to minimize:

$$\text{(Eq. 10-18)} \qquad \mathrm{E_{GRAPH\_ILLUMINATION}} = \int_{\Omega} (\mathrm{Gl}(\mathrm{n} + \mathbf{V}(\mathrm{n})) - \mathrm{Gl}(\mathrm{n}))^2 d\Omega$$

This energy shows the illumination change of node **n** through motion V and we demand that this change has to be minimum. With (Eq. 10-17) we were able to explain how the graph needs to deforms by external forces: spatial ($\int_{\Omega} ((\mathrm{G_{t-1}} + \mathbf{V}(\mathrm{n})) - \mathrm{G_t})^2 d\Omega$) and illumination ($\mathrm{E_{GRAPH\_ILLUMINATION}}$).

We know ([121][122][123]) that for objects that undertake a non-rigid

motion under varying illumination we can write:

$$\text{(Eq. 10-19)} \qquad \mathrm{F_t}(\mathbf{X}) = \mathrm{F_{t-1}}(\mathbf{X}) + \boldsymbol{l}_{t-1} \cdot \boldsymbol{b}_d + \Delta\boldsymbol{l} \cdot \boldsymbol{b}_{t-1}$$

where $\boldsymbol{l}_{t-1}$ are the illumination coefficients of t-1 frame, $\boldsymbol{b}_d$ the deformation image basis because of deformation of object shape, $\Delta\boldsymbol{l}$ is the illumination variation due to change of direction and intensity of illumination and $\boldsymbol{b}_{t-1}$ the image basis of the previous frame. We have showed in (Eq. 10-14) and (Eq. 10-15) that $\Delta\boldsymbol{l}$ changes due to rigid motion $\mathbf{V}(n)$ and it is minimized inside $E_{GRAPH\_DEFORMATION}$ (Eq. 10-17). Therefore the last term of (Eq. 10-19) is neglected. Finally, $\boldsymbol{b}_d$ explains the local deformation of the object and using (Eq. 10-14) it is given by:

$$\text{(Eq. 10-20)} \qquad \mathbf{b_d} = \rho\mathbf{r}\mathbf{Y}(\mathbf{N} + \boldsymbol{\Delta}\mathbf{N})\mathbf{U} = \rho\mathbf{r}\boldsymbol{\nabla}_{\mathbf{N}}(\mathbf{N})\boldsymbol{\nabla}_{\mathbf{x}}\mathbf{N}\,\mathbf{U} = \mathbf{A}\,\mathbf{U}$$

Note that **U** is the decomposition of the local displacement **u** through FEM, B-Splines, etc.

Using (Eq. 10-9), (Eq. 10-13), (Eq. 10-14), (Eq. 10-15), (Eq. 10-16), (Eq. 10-17), (Eq. 10-18), (Eq. 10-19) and (Eq. 10-20) we can calculate the energy to minimize in case of a non-rigid motion with deformable objects under varying illumination as:

$$\text{(Eq. 10-21)} \qquad E_{MUTUAL_{ILLUMINATION}} = a * E_{GRAPH_{DEFORMATION}} + (1-a) * \sum_{i=1}^{Ni} E_{OBJECT_{DEFORMATION}}$$

$$E_{MUTUAL\_ILLUMINATION} = a * E_{GRAPH_{DEFORMATION}} + (1-\mathrm{a}) * \sum_{\mathrm{i}=1}^{\mathrm{Ni}} \left[\int_{\Omega i} \sigma_{node}^{t} \varepsilon_{node} d\Omega_i + \int_{\Omega i} (\mathrm{F_t}(\mathrm{x} + \boldsymbol{V}(n) - \boldsymbol{U}(\mathrm{x})) - \mathrm{F_{t-1}}(\mathrm{x}))^2 d\Omega_i \right] \Leftrightarrow$$

$$E_{MUTUAL\_ILLUMINATION} = a * E_{GRAPH_{DEFORMATION}} + (1-\mathrm{a}) * \sum_{\mathrm{i}=1}^{\mathrm{Ni}} \left[\int_{\Omega i} \sigma_{node}^{t} \varepsilon_{node} d\Omega_i + \int_{\Omega i} (\mathrm{F_{t-1}}(\mathrm{x}) + \boldsymbol{l}_{t-1} \cdot \boldsymbol{b}_d - \mathrm{F_{t-1}}(\mathrm{x}))^2 d\Omega_i \right]$$

$$= a * E_{GRAPH_{DEFORMATION}} + (1-\mathrm{a}) * \sum_{\mathrm{i}=1}^{\mathrm{Ni}} \left[\int_{\Omega i} \sigma_{node}^{t} \varepsilon_{node} d\Omega_i + \int_{\Omega i} (\boldsymbol{l}_{t-1} \mathbf{A}\, \mathbf{U})^2 d\Omega_i \right]$$

The solution can be estimated using PCA analysis (see [123] for a detailed solution) or alternate minimization (see [122]). We have not included the term $\int_{\Omega i} |\nabla \mathrm{F_t}(\mathrm{x} + \boldsymbol{V}(n) - \boldsymbol{U}(\mathrm{x})) - \nabla \mathrm{F_{t-1}}(\mathrm{x})|^2 d\Omega_i$ of (Eq. 10-12) since small variations of intensity are explained with illumination change. The importance of (Eq. 10-21) lies on its ability to unify the deformation of the graph, the deformation of the nodes (objects) and the deformation of illumination through an energy formulation. Additionally, separating the last two terms of (Eq. 10-19) we are able to deal with $\Delta \boldsymbol{l}$ and $\mathbf{b}_\mathrm{d}$ independently and explain: a) the deformation of the object ($\int_{\Omega i} \sigma_{node}^{t} \varepsilon_{node} d\Omega_i$), the change of intensity through deformation ($\int_{\Omega i} (\boldsymbol{l}_{t-1} \mathbf{A}\, \mathbf{U})^2 d\Omega_i$) and the change of intensity through rigid motion ($\int_{\Omega} (\mathrm{Gl}(\mathrm{n} + \mathbf{V}(\mathrm{n})) - \mathrm{Gl}(\mathrm{n}))^2 \mathrm{d}\Omega$).

Finally in a similar way we can rewrite the minimization of the error in case of non mutual information as:

(Eq. 10-22) $E_{NON\_MUTUAL} =$

$$\int_{\Omega'} (\mathrm{F_t}(\mathrm{x} - \boldsymbol{U}(\mathrm{x})) - \mathrm{F_{t-1}}(\mathrm{x}))^2 d\Omega' - \int_{\Omega'} \frac{<U(x), V(x)_{INTERPOLATED})>}{\|U(x)\| \|V(x)_{INTERPOLATED}\|} d\Omega'$$

$$E_{NON\_MUTUAL} = \int_{\Omega'} (\boldsymbol{l}_{t-1}\mathbf{A}\,\mathbf{U} + \Delta\boldsymbol{l} \cdot \boldsymbol{b}_{t-1})^2 d\Omega' - \int_{\Omega'} \frac{< \mathbf{U}, \mathbf{V}_{INTERPOLATED}) >}{\|\mathbf{U}\| \|\mathbf{V}_{INTERPOLATED}\|} d\Omega'$$

Again $\mathbf{U}$, $\Delta\mathbf{l}$ can be estimated by alternate minimization between $\mathbf{U}$, $\Delta\mathbf{l}$ as described in [122] and $\mathbf{V}_{INTERPOLATED}$ is calculated with (Eq. 10-17).

10.8 Results

In this section we present a graphic result of our interpolation methodology given frames A and I (B-C intermediate created frames).

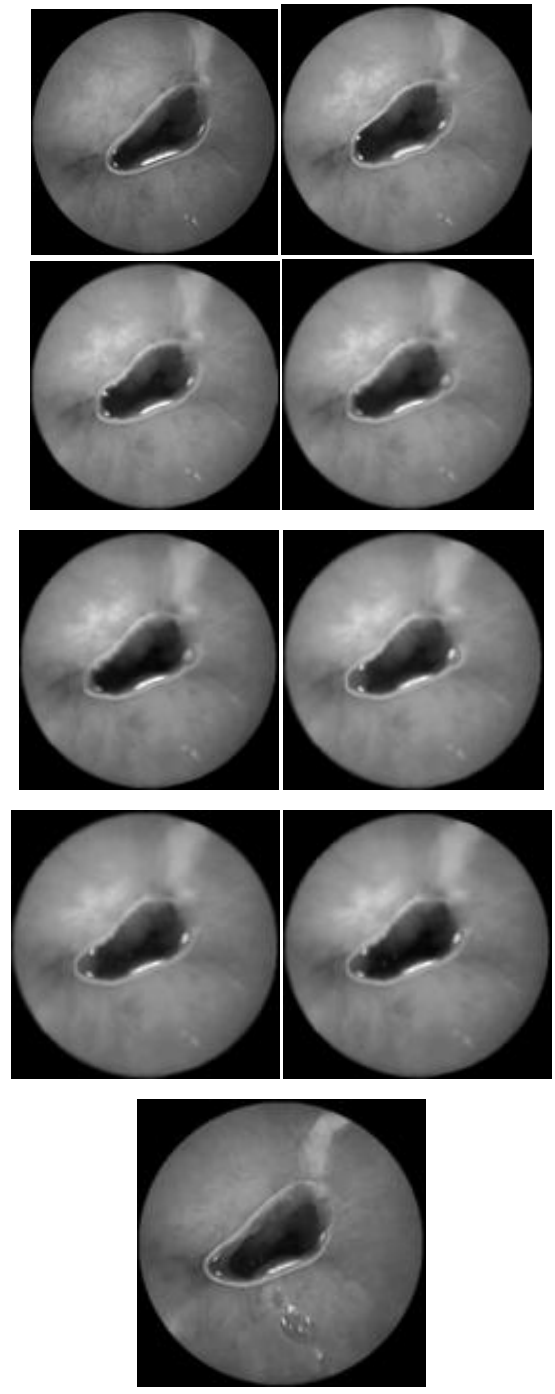


Fig. 10-5. A graphic example of the proposed elastic video interpolation methodology

In this example we have set the methodology to produce 7 interpolated frames. Elasticity of the graph was determined by Young's elasticity modulus

with $E=10^4$ and Poisson's ratio v= 0.2. Since no study was found in the literature that can give the elastic properties of small intestine, we figured out these values by trial and error so that the graph and its nodes are so elastic that explain a smooth transition without "breaking" the edges of the objects.

10.9 Conclusions

In this chapter we have presented a novel video interpolation methodology for a challenging medical problem. This new framework managed to give us very promising results and currently we are testing its usability in other cases, such MRI and CT modalities.

The novelties of the framework can be synopsized in the following points:

1) The graph G which describes the structure and connectivity of the objects of the image was assigned with elastic properties. These elastic properties explain the motion and displacement of the camera and objects.
2) Using mathematical formulations we were able to describe through energy perspective how the graph can deform to release amounts of energy to move and deform regions (spatial variation).
3) The graph is extended to the pixel level through its nodes that describe regions. Nodes also carry elastic properties that can explain the local deformations. In this way the energy of the image transverses from top (graph) down to the bottom (pixel).

4) We were able to describe how energy of the graph can explain not only the motion and deformation of the regions but also how the illumination changes ("color" variation).

Creating interpolated frames we can now use the concepts of [119] to create a detailed 3D reconstruction of the viewable digestive tract. The proposed 3D scheme [119] has been well approved by our collaborating gastroenterologists as it is a powerful enhancement to existing capsule endoscopy technology.

Apart from this important feature (3D reconstruction) we want to extend our framework to deal with more complicated cases such as shown in Fig. 10-6 . Using super-resolution image reconstruction techniques, depth estimation and graph interpolation we want to create more robust interpolated frames in case of huge motions of the camera.

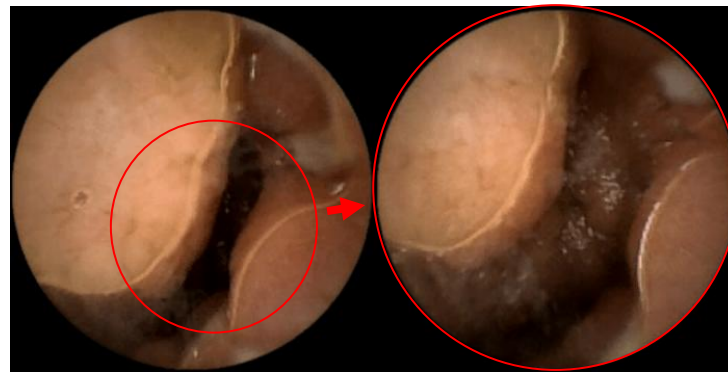


Fig. 10-6. Two consecutive frames where the camera moves through the intestine forward. More information is revealed from a smaller area of the original frame

Additionally we want to extend the graph capabilities so that it is able to add/remove nodes through energy formulations.

CHAPTER 11: CONCLUSIONS AND RESEARCH CONTRIBUTIONS

11.1 Synergistic Methodology

The novelty of this dissertation lies on the creation of an intelligent diagnosis methodology for gastroenterologists. The uniqueness of the system can be synopsized in the approach to deal with the WCE problems: **A wide range of specialized methodologies addressing each WCE issue individually and altogether working synergistically to produce the final result.**

An essential contribution of the methodology is that it is not limited only to WCE cases but **it can also be used for the majority of 2D medical images**. Because of the synergistic design of the methodology the scheme can be **easily extended** to support other processing features through the **expansion** of its low level layer (see Fig. 3-1).

Finally, the Graphic User Interface (GUI) of the system offers the ability to physicians to interact with the system and perform various medical imaging processes.

The following full-paper refereed publications support the new scientific contributions of the CAD system:

Journals

1) A. Karargyris, N. Bourbakis, "A Survey on Wireless Capsule Endoscopy and Endoscopic Imaging", IEEE Engineering in Medicine and Biology Society (IEEE

MBS Magazine January-February 2010).

2) A. Karargyris, N. Bourbakis, "Three-dimensional reconstruction of capsule endoscopy videos", (to be submitted IEEE Transactions on Medical Imaging)

3) A. Karargyris, N. Bourbakis, "Enhancement of detection of blood-based abnormalities using registration-like methodology with statistical comparisons", (to be submitted at IEEE Transaction on Information Technology in Biomedicine)

4) A. Karargyris, N. Bourbakis, "Data Mining techniques on polyps and ulcer in capsule endoscopy videos using registration", (to be submitted IEEE Transactions on Medical Imaging)

Conferences

5) A. Karargyris, N. Bourbakis "An Elastic Video Interpolation Methodology in Wireless Capsule Endoscopy videos", IEEE International Conference on "Bioinformatics and Bioengineering (BIBE-2010)", May 31st – June 3rd 2010, Philadelphia, USA

6) A. Karargyris, O. Karargyris, N. Bourbakis, "3D representation of the digestive tract surface in Wireless Capsule Endoscopy videos", ", IEEE International Conference on "Bioinformatics and Bioengineering (BIBE-2010)", May 31st – June 3rd 2010, Philadelphia, USA

7) Karargyris, A.; Bourbakis, N., "A data mining algorithmic approach for processing wireless capsule endoscopy data sets," *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE* ,

vol., no., pp.6636-6639, 3-6 Sept. 2009

8) Karargyris, A.; Bourbakis, N., "A video-frame based registration using segmentation and graph connectivity for Wireless Capsule Endoscopy," *Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH* , vol., no., pp.74-79, 9-10 April 2009

9) Karargyris, A.; Bourbakis, N., "Identification of polyps in Wireless Capsule Endoscopy videos using Log Gabor filters," *Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH* , vol., no., pp.143-147, 9-10 April 2009

10) Karargyris, A.; Bourbakis, N., "A methodology for detecting blood-based abnormalities in Wireless Capsule Endoscopy videos," *BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on* , vol., no., pp.1-6, 8-10 Oct. 2008

11) Karargyris, A.; Bourbakis, N., "Identification of ulcers in Wireless Capsule Endoscopy videos," *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on* , vol., no., pp.554-557, June 28 2009-July 1 2009

11.2 Analysis of the new research contributions

The new research contributions can be synopsized through the following five (5) points:

11.2.1 Registration scheme

A new registration methodology has been created to deal with the reg-

istration of WCE frames. In chapter 4 it was shown that WCE frames exhibit a tremendously dynamic variation. Traditional and modern registration methods did not deal with the problem successfully. Thus, we designed a registration scheme to perform the process by taking into account how gastroenterologists perform registrations. This helped utilize the graphs to describe regions and objects in a pyramid fashion. This structural approach of image interpretation led to an algorithm which is more human-perceived than simply statistical (i.e. texture, color, machine learning).

11.2.2 Blood-based abnormalities

Identification of blood-based abnormalities is a difficult problem in WCE images since they exhibit similar textural and color characteristics with normal digestive tract tissues. We approached this problem in a non-traditional fashion (a- color and texture descriptors extraction and b- neural network classification).We focused on distinguishing the color of the blood-based patterns from other tissues by creating a different color space where the new color components are as much uncorrelated as possible. This important notion is also supported by later works such as [124]. This uncorrelation characteristic was able to give us satisfying results. Extension of this algorithm with the traditional approach (texture and classifiers) could produce better results. Additionally, our work showed that blood-based abnormalities consist of three (3) different types, each one with different characteristics. Therefore, it is important to take this feature into consideration when dealing with the blood-based cases.

11.2.3 Extraction of polyps

Detection of polyps in WCE images is also a challenging and complex problem as shown in chapter 6 in detail. Our research was the first that addressed this issue on an exclusive basis. Whereas other authors dealt with this issue by using color, texture and classification methods, our methodology utilized the most important characteristic of polyps: shape. And that is because polyps vary much in color and texture. Even without the use of classifiers this methodology has reached high levels of sensitivity and specificity.

11.2.4 Data mining

Apart from developing new algorithms that address various WCE issues, our research was the first to perform data mining on WCE images and it achieved to extract hidden knowledge between ulcers, polyps and healthy tissues. Although it would be expected that no relationships between the three (3) categories could be found, after applying data mining techniques important information has been extracted and it is discussed in detail in 8.2.1 and 8.2.2.

11.2.5 3D reconstruction of digestive tract

Finally, a novel methodology was developed to create a 3D reconstruction of the digestive tract based on WCE frames. The use of such technique is really a contribution since it offers an improved view of the digestive tract for

gastroenterologists. The novelty also lies on the fact that this algorithm works with no change of the current WCE technology.

To further improve the 3D reconstruction scheme we developed our own video interpolation scheme that is able to create intermediate frames from given WCE frames. The novelty of this methodology consists of a) adding elastical features to the global graph of the image, and b) importing the phenomenon of peristalsis through energy formulations that explain the deformation of the graph and the intensity change simultaneously.

11.3 Further and Future Research

With this dissertation a scheme was proposed to address the detection of abnormalities in WCE videos as well as the reduction of examination time. Because of the diversity of the abnormalities, the design of the system's architecture was carefully planned to have the ability to be extended to address different cases. Overall the system managed to deal with these problems successfully. However as in any image processing case there is always room for improvement.

More specifically regarding the detection of polyps an enhancement would be to use the proposed methodology in conjunction with the 3D reconstruction scheme. This means that the 2D shape of the potential polyps (protrusions) could be extended to 3D. Application of curvature features on the extracted 3D shapes could increase the detection rate. Some additional geometric characteristics (i.e. roughness of polyps' borders) could further improve the

overall algorithm.

In chapters 8 and 9 an improvement to existing WCE technology was shown. Although the 3D reconstruction offers impressive results and it offers the ability to the user to view WCE frames in a different way, aspects of future research have been pointed out. More specifically, current technology of sensors can offer the ability to view digestive tract in real 3D.

Due to the research carried throughout the dissertation we were able to design advanced capsules for diagnostic and therapeutic purposes. A detailed description is given later in the appendix.

APPENDIX A: DEVELOPMENT OF GRAPHIC USER INTERFACE TOOL

ATRC Video Toolbox (ATRC-VT) is the software that has been developed in our lab to carry out with the detection of abnormalities and it utilizes all the aforementioned methodologies.

The purpose of developing our software with graphic user interface was clearly for simplicity since it is aimed for use by beginner users, such as physicians and nurse personnel.

In this chapter we will give information on the structure of the software.

12.1 Software Layout

In Fig. 12-1 a screenshot of the toolbox is given. The software was designed to perform various tasks besides running the detection methodologies as described above. Therefore, the software offers image processing techniques used by our methodologies:

1) Color transformation to numerous color spaces
2) Segmentation steps
3) Wavelet transformation

We are still in the process of expanding the software to include settings for a fuzzy neural network and the choice to select which abnormality the user would like to detect.

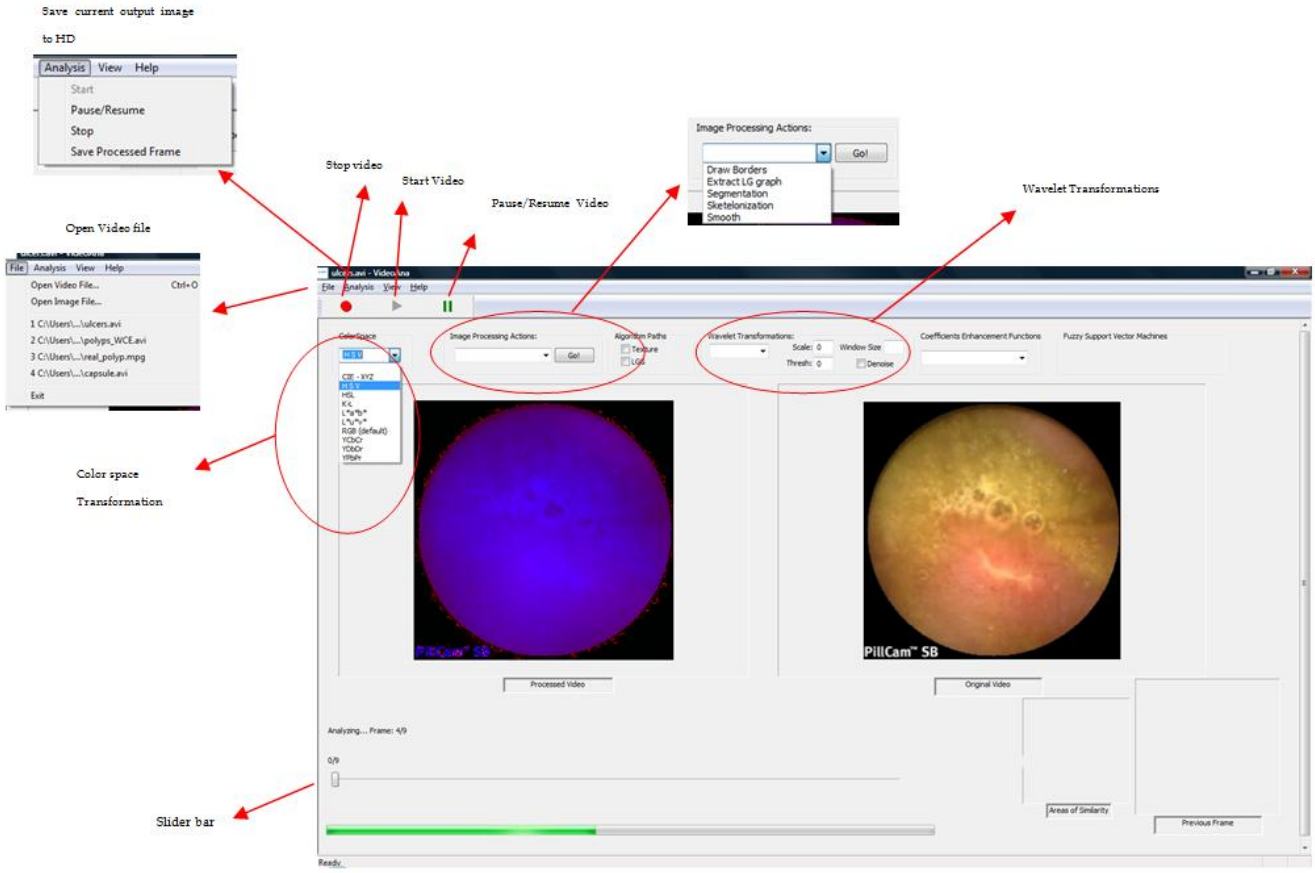


Fig. 12-1. Snapshot of ATRC-VT with explanations

12.2 Software Engine

The engine is based on Microsoft DirectShow multimedia framework. Directshow provides the necessary infrastructure (development libraries/APIs) to handle multimedia files such as video and audio. In our case the engine is pretty much straightforward:

a) Source filter to read raw data from the input video file.

b) Sample Grabber filter to retrieve samples as they pass through the filter graph.

c) Render filter to display the input frame as well as the processed frame on the screen.

The most important aspect of the engine is b) which consists of the ISample-Grabber() function which is a Directshow filter. ISampleGrabber() was set in SetCallback mode (method) so that a specified callback function, BufferCB(), which points to the current sample buffer. After that, the function calls the main processing function mainFunction() where the data of the buffer are handled as regular BMP file. Inside mainFunction() all the described methodologies are called through methods. The output is saved on a different buffer and both the input buffer and the processed buffer are displayed on the two boxes of the program (Fig. 12-1).

APPENDIX B: DESIGN OF ADVANCED CAPSULES FOR ENDOSCOPY

13.1 Design of a new wireless capsule endoscopy

"Researchers [132] have proposed the idea that endoscopists will be able to 'control and steer' the Capsule Endoscopy (CE), as they are able to do in standard endoscopy. This would mean being able to maintain the capsule steady in a selected area and hold the view in order to have more time to examine the opposite wall of the bowel".

Based on these thoughts several designs of the endoscopic capsule have been proposed. Some of them, [133]-[136], are shown below:

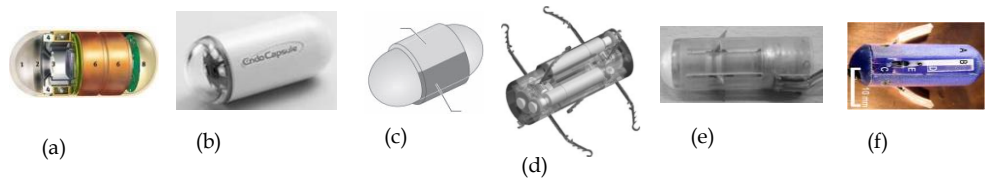


Fig. 13-1. Various capsule prototypes from different sources and groups

These are all very good ideas with various potentials in each design. For example, the capsule marketed by Given Imaging Inc. is simple with limited capabilities but with better safety from other designs in case of blockages in the intestine. The capsule from Olympus Corp. has more capabilities due to its magnetic fields for possible forward and backward motion guided by the gastroenterologist. Prototypes Fig. 13-1-d and Fig. 13-1-e present capabilities of

forward and backward motion due to their robotic legs. However, these two capsule designs have low safety features due to high risk of a possible grabbing problem inside the intestine wall and at the same time they may damage the wall causing bleeding. The last capsule (Fig. 13-1-f) offers better grasping and smooth legs minimizing bleeding accidents.

Our research efforts proposed here address the detection of abnormalities in WCE videos as well as the reduction of examination time. Because of the diversity of the abnormalities, the design of the system's architecture was carefully planned to address different cases.

Using the knowledge we have extracted throughout this research we can tell how an ideal capsule endoscopy for diagnostic purposes can be constructed:

-Depth cameras (Time-of-Flight cameras) have reached a mature level of resolution and detail. The problem is the size of these cameras which in just a few years it has been reduced to the size of 1.5 inch. Given current trend we are sure that the miniaturization of such camera will soon be achieved.

-Accelerometers and magnetometers offer direction and acceleration information. Fortunately these sensors are so small that their size is less than half of a fingertip.

-Narrow band light filters can be utilized in existing cameras to reveal additional information on the surface of the digestive tract. Such sensors are used in other endoscopy devices.

-Existing miniature cameras offer a viewing field of more 100^{o} degrees.

Use of such cameras could increase the existing viewing field. Use of two (2) cameras sitting on the opposite sides of the capsule could tremendously increase the viewing field. Ideally, if cameras achieved nearly 180^{o} degrees viewing field, then using two (2) of those would reach a 360^{o} degrees viewing field.

-Power consumption is probably the most important issue in capsule endoscopy. Some researchers have proposed use of external magnetic field to power the capsule. However studies have to be carried out to examine health risks. We believe that redesigning the way current technology transmits information to the patient's receiver is a potential solution: Since the capsule constantly transmits video frames to the receiver a lot of energy is wasted. A different approach would be that the capsule stores all the information on a memory bank and by the end of the procedure the video could be uploaded to a workstation. In this way even current battery technologies could be used to increase the resolution of the video frame and more importantly increase the frame rate which currently is really low.

Thus, it is a challenge for us to propose a new design, doable with as many features from the discussion above [138]. Having thoroughly examined the problem of WCE we focused our research on two goals: the design of a) a diagnostic capsule and b) a robotic therapeutic capsule. In this way we managed to combine all the state-of-art features in each capsule to serve their purposes with maximum efficiency. Additionally, characteristics that are not needed on either capsule were excluded to reduce the cost. Finally both designs are as safe as the Given Imaging and Olympus capsules are.

Synopsizing, after discussion with gastroenterologists, the diagnostic capsule should be able to:

1. Capture video at much larger video frame,
2. Have the potential of grabbing 3D information of the gastric structures,
3. Support with position information while traversing inside the intestine.

Having these features we propose the following diagnostic capsule:

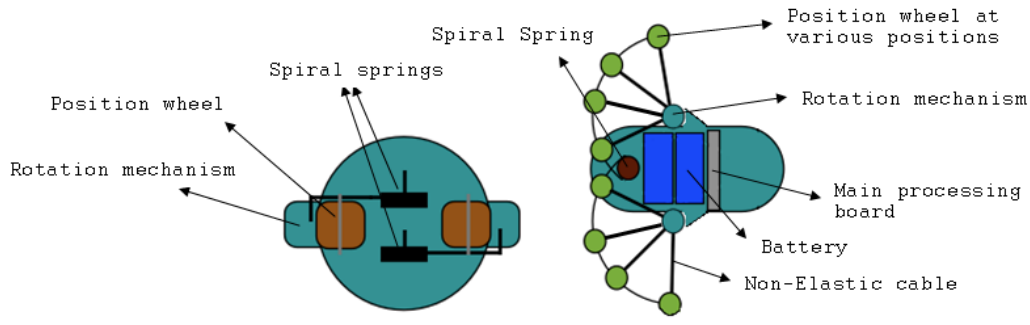


Fig. 13-2. Diagnostic Capsule. Left: Face of the capsule with 2 rubber - position sensors wheels. A spiral spring expands/retracts each wheel. Right: Profile of the capsule showing the expansion in various positions of the wheels

Each position sensor is attached to a soft spiral spring that has the ability to expand as much as possible to reach the intestinal wall (this depends of how much the intestine is deformed) and let the precision sensor rotate due to frictional forces. More specifically, the two (2) position sensors basically follow the same principal as odometers. They are composed of frictional rubber materials so that when they are in contact with the soft tissue of the intestinal wall they are able rotate. This rotation gives information about how far the capsule

has moved inside the intestine by emitting electromagnetic peaks that are received by precision sensors on the control board of the capsule.

The two (2) position sensors offer not only accurate and robust information (in comparison with a single position sensor) about the position of the capsule but also work as guidance for the capsule to keep as steady as possible on its traversing direction. Finally, due to its expansion capabilities the capsule can successfully record video from the large bowel whereas with current capsule technology its small size does not allow for a robust nice video of the large bowel.

We are currently experimenting on a) the type of the springs and b) the material of the rubber wheels to be used for the aforementioned purposes. Such materials exist and serve these goals. By knowing the position of the captured frame inside the intestine, gastroenterologists will be able to know where a potential abnormal case is inside the gastric tract. This information is extremely important for future reference or for a possible surgery. By knowing the position of the captured frame inside the intestine, gastroenterologists will be able to know where a potential abnormal case is inside the gastric tract. This information is extremely important for future reference or for a possible surgery.

It must be mentioned that the proposed capsule has been carefully planned to have the same size as the current commercial capsules (27mmX13mm). Additionally, the capsule does not require a wearable receiver since it does not transmit the video wirelessly but on the contrary it saves the

video on memory storage. After the process ends, the front part of the capsule, which contains the memory storage, can be easily removed and connected to a memory reader to upload the video to a workstation.

Whereas the proposed diagnostic capsule is a passive device the proposed therapeutic capsule is an active fully controllable robotic device. Additionally, the capsule barely exceeds current commercial capsules' size. In Fig. 13-3 to Fig. 13-5 the conceptual representation of the proposed robotic camera is given.

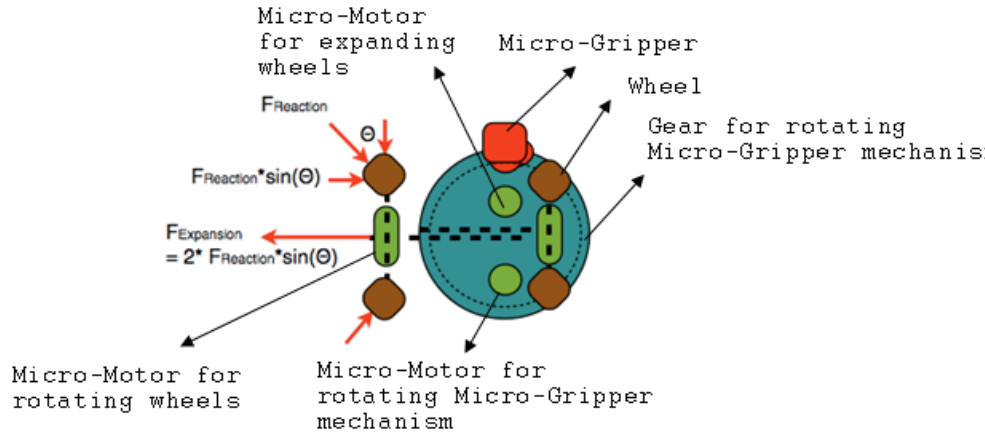


Fig. 13-3. Face view of the robotic capsule with one wing expanded. Note that the wings are fully retracted inside the body of the capsule to lose contact with the gastric tract. Applied forces are also drawn.

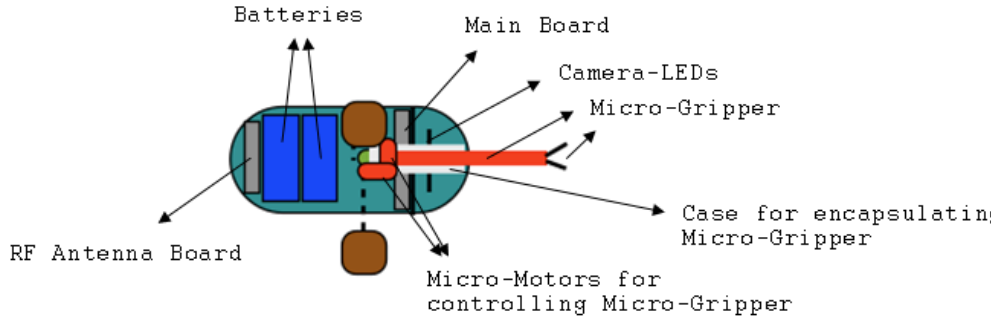


Fig. 13-4. Top view of the robotic capsule with one wing expanded. Micro-

Gripper is shown here expanded.

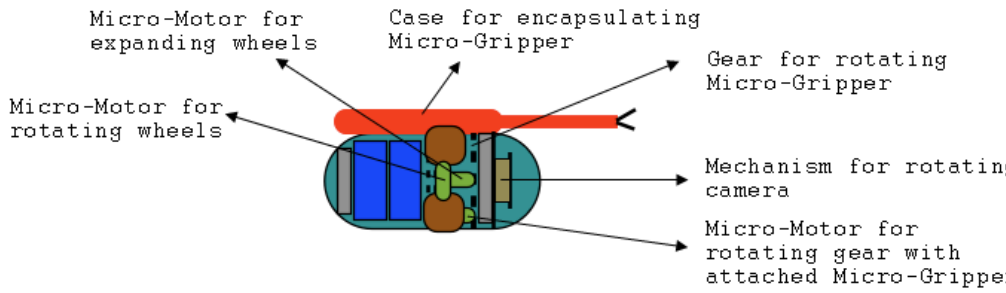


Fig. 13-5. Side view of the robotic capsule. The capsule is barely larger than commercial capsule due to its Micro-Gripper.

As seen from Fig. 13-3 to Fig. 13-5 the capsule consists of an advanced and complex mechanism that:

1. Expands the wings that carry four (4) wheels to hold the capsule at the current position in the intestine.
2. Offers the ability to expand size of the capsule's width three (3) times. This enables the capsule to fully function its capabilities in the large bowel too.
3. Offers the ability to rotate the wheels both directions and therefore move the capsule forwards/backwards.
4. Ability to carry a Micro-Gripper to take biopsy of the intestine. The gripper can be replaced with a needle or a cauterizing device.
5. Ability to rotate the Micro-Gripper to some degree of angle to have a better access area.

In the next section physical principals are explained in detail about the robotic capsule.

13.2 Capsule's Functional Specifications and Features

In this sub-section we briefly present the specs of our design and the principle of physics associated with the forces applied on the capsule mainly due to peristalsis (P), gravity (G) and friction (F), Fig. 13-6.

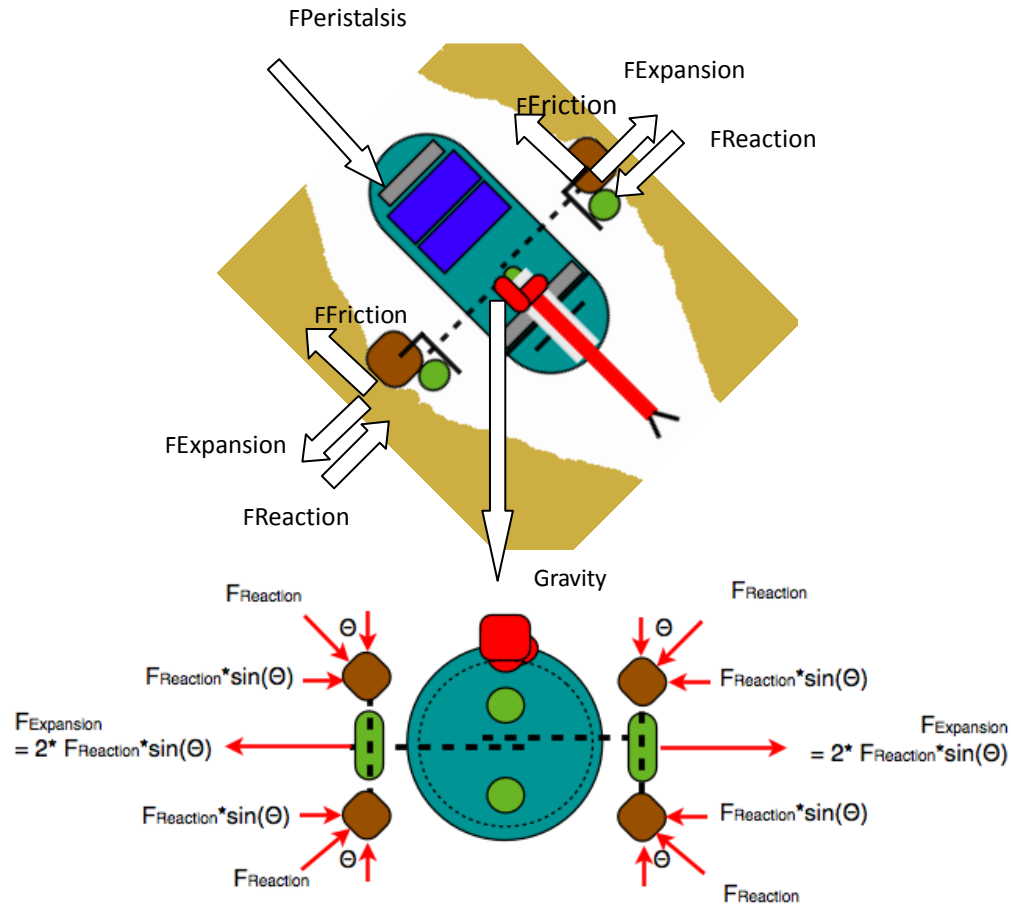


Fig. 13-6. Two views (top: profile, bottom: face) of the capsule with the forces applied to it are shown.

From the Fig. 13-6 we can calculate the forces applied to the capsule in

case of vertical position where the gravity force is at its maximum:

(Eq. 13-1) Total Expansion force = $4*F_{Reaction} * \sin(\Theta) = F_{Peristalsis} + m*g$

For each wheel:

(Eq. 13-2) $F_{Friction} = \mu*F_{Reaction}$

μ : the friction coefficient of the wheel material

For each motor to rotate two (2) wheels:

(Eq. 13-3) Torque = $2*\mu*F_{Reaction}$*Wheel_Radius

Wheel_Radius: 2.5mm

The physical sequence of the forces is: $F_{Peristalsis}$ + Gravity -> $F_{Expansion}$ -> $F_{Reaction}$ -> $F_{Friction}$ -> Motor Torque

For a 27mmX13mm and 50-gram capsule the $F_{Peristalsis}$ is approximately at 0.5N [137] and the gravity force is 0.5N.

From (Eq. 13-1) the total expansion force is 1N and the FReaction is 353mN. In that case from (Eq. 13-3) the torque must be $2*353*0.0025*\mu = 1.765*\mu$ mNm

In other words, in case the capsule has to halt at a certain position, the wings expand to neutralize the peristalsis force and the gravity force applied to

the capsule. The expanded wings will touch and deform the intestinal wall through their attached rubber wheels. Consequently, this will generate reaction forces from the intestine to the capsule and these reaction forces will cause friction forces on the wheels. Finally the motor will have to provide with sufficient torque to overcome the torques of the friction forces. There are miniature motors that have this capability. A quick research in the miniature motor industry (i.e. Namiki Co., New Scale Technologies) showed that there are multiple micro motor models that offer sufficient power, torque and size capabilities. Careful planning must be taken to find materials that provide the right friction coefficient which can halt the capsule and allow the motors' torque to rotate them. After examining the literature (i.e. [136]) we are confident that these materials exist. Additionally, the wheels can be designed to have soft rubber cut teeth to enable the better rotating momentum.

REFERENCES

[1] Yu M. M2A capsule endoscopy. A breakthrough diagnostic tool for small intestine imaging. Gastroenterol Nurs. 2002; 25:24-27.

[2] Minimally Invasive Therapy & Allied Technologies Apr 2007, Vol. 16 Issue 2, p91-100

[3] Vilarino, F. Spyridonos, P. Pujol, O. Vitria, J. Radeva, P. Universitat Autonoma de Barcelona, Spain, Automatic Detection of intestinal juices in wireless capsule video endoscopy, Proceedings of the 18th International Conference on Pattern Recognition - Volume 04, Pages: 719 - 722 , 2006.

[4] Kodogiannis, V.S. Boulougoura, M., Images Sch. of Comput. Sci., Westminster Univ., London, UK, Neural Networks-Based Approach for the Classification of Wireless-Capsule Endoscopic, Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference, 31 July- 4 Aug.2005 Volume: 4, Page(s): 2423- 2428 vol. 4

[5] Kodogiannis, VS (2004). Computer-aided diagnosis in clinical endoscopy using neuro-fuzzy systems. In Proc's IEEE Int. Conf. on Fuzzy Systems, Vol. 3, pp. 1425-29.

[6] Kodogiannis, VS & M. Boulougoura (2007). An adaptive neurofuzzy approach for the diagnosis in wireless capsule endoscopy imaging. In Int. J. of Inf. Technology, Vol. 13, pp.46-56.

[7] Szczypinski, P., Sriram, R.D., Sriram, P., Reddy, D., Model of Deformable Rings for Aiding the Wireless Capsule Endoscopy Video Interpretation and Reporting, Proceedings of the International Conference on Computer Vision and Graphics 2004, Warsaw, Poland, September 22-24, 2004.

[8] JeongKyu Lee, JungHwan Oh, Xiaohui Yuan, and Shou-jiang Tang. "Automatic Classification of Digestive Organs in Wireless Capsule Endoscopy Videos". Proc. of the 22nd Annual ACM Symposium on Applied Computing 2007. Seoul, Korea, March 11 - 15, 2007.

[9] Sae Hwang, JungHwan Oh, Jay Cox, Shou-jiang Tang, and Harry F. Tibbals. "Blood Detection in Wireless Capsule Endoscopy using Expectation Maximization Clustering". Proc. of SPIE International Symposium, Medical Imaging 2006. 11 - 16 February 2006. San Diego, California

[10] D. Rex, R. Weddle, D. Pound, K. O'Connor, R. Hawes, R. Dittus, J. Lappas, and L. Lumeng, "Flexible sigmoidoscopy plus air contrast barium enema versus colonoscopy for suspected lower gastrointestinal bleeding," Gastroenterology, vol. 98, pp. 855–861, 1990.

[11] The Expectation Maximization Algorithm: A short tutorial. Sean Borman, July 18 2004, Last updated October 14, 2006 (http://www.seanborman.com/publications/EM_algorithm.pdf)

[12] Yair Weiss. Bayesian motion estimation and segmentation. PhD thesis, Massachusetts Institute of Technology, May 1998.

[13] Berens, M. Mackiewicz, M. Fisher, and G. D. Bell, "Using color distributions to discriminate tissues in Wireless Capsule Endoscopy images.," in Proceedings of Medical Image Understanding and Analyses 2005 Conference, Bristol, UK, July 2005.

[14] M. Mackiewicz J. Berens M. Fisher D. Bell, Color and Texture Based Gastrointestinal Tissue Discrimination, Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, Volume: 2

[15] M. Coimbra P. Campos J.P.S. Cunha, Topographic Segmentation and Transit Time Estimation for Endoscopic Capsule Exams IEETA, Department of Electronics and Telecommunications, University of Aveiro. miguel.coimbra@ieeta.pt; Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, Volume: 2.

[16] C.J Burges, "A tutorial on support vector machines for pattern recognition", in Knowledge Discovery Data Mining, vol.2, no.2, 1998, pp. 1-43.

[17] International organization for standardization – Organisation Internationale de Normalisation ISO/IEC JTC1/SC29/WG11 coding of moving pictures and audio SO/IEC JTC1/SC29/WG11N6828 Palma de Mallorca, October 2004

[18] Magoulas G: Neuronal networks and textural descriptors for automated tissue classification in endoscopy. Oncol Rep 15:997-1000, 2006.

[19] Magoulas, G., Plagianakos, V. and Vrahatis, M., Neural network-based colonoscopic diagnosis using on-line learning and differential evolution. Appl. Soft Comput. v4. 369-379.

[20] R. Storn, K. Price, Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (1997) 341-359.

[21] Bourbakis, N.; Makrogiannis, S.; Kavraki, D., A neural network-based detection of bleeding in sequences of WCE images, Bioinformatics and Bioengineering, 2005. BIBE 2005. Fifth IEEE Symposium on, Volume , Issue , 19-21 Oct. 2005 Page(s): 324 – 327

[22] Karkanis, S.A.; Iakovidis, D.K.; Maroulis, D.E.; Karras, D.A.; Tzivras,Computer-aided tumor detection in endoscopic video using color wavelet features,M.Information Technology in Biomedicine, IEEE Transactions on Volume 7, Issue 3, Sept. 2003 Page(s): 141 – 152

[23] G. Van de Wouver, P. Scheunders, S. Livens, and D. Van Dyck, "Wavelet correlation signatures for color texture characterization," Pattern Recognition, vol. 32. pp. 443-451, 1999

[24] Maroulis D.E.1; Iakovidis D.K.; Karkanis S.A.; Karras D.A. , CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames Source: Computer Methods and Programs in Biomedicine, Volume 70, Number 2, February 2003, pp. 151-166(16)

[25] Wang, P. Krishnan, S.M. Kugean, C. Tjoa, Classification of endoscopic images based on texture and neural network, Wang, P. Krishnan, S.M. Kugean, C. Tjoa, M.P. Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE, Volume: 4, On page(s): 3691- 3695 vol.4.

[26] Karargyris, Alexandros; Bourbakis, Nikolaos, "A methodology for detecting blood-based abnormalities in Wireless Capsule Endoscopy videos," BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on , vol., no., pp.1-6, 8-10 Oct. 2008

[27] Y. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation," Computer Graphics and Image Processing, Vol. 13, No. 3, July, 1980, pp. 222 - 241.

[28] A.Moghaddamzadeh and N.Bourbakis, A fuzzy-like region growing approach for segmentation of colored images, PR Society J. Pattern Recognition, vol. 30,6, 867-881,1997

[29] Bourbakis, N., Detecting abnormal patterns in WCE images, 5th International IEEE Bioinformatics and Bioengineering, 2005. BIBE-05, 19-21 Oct. 2005 Page(s): 232 – 238

[30] Bourbakis, N.; Makrogiannis, S.; Kavraki, D., A neural network-based detection of bleeding in sequences of WCE images, Bioinformatics and Bioengineering, 2005. BIBE 2005. Fifth IEEE Symposium on, Volume , Issue , 19-21 Oct. 2005 Page(s): 324 – 327

[31] G. Van de Wouver, P. Scheunders, S. Livens, and D. Van Dyck, "Wavelet correlation signatures for color texture characterization," Pattern Recognition, vol. 32. pp. 443-451, 1999.

[32] T. P. Weldon, W. E. Higgins, and D. F. Dunn, ``Gabor filter design for multiple texture segmentation,'' Optical Engineering, vol. 35, no. 10, pp. 2852-2863, Oct. 1996.

[33]] J. Daugman. Uncertainty relation for resolution in space, spatial frequency and orientation optimized by twodimensional visual cortical filters. J. Opt. Soc. Am., 2(7):1160–1169, 1985.

[34] M. A. Webster and R. L. De Valois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," J. Opt. Soc. Am. A 2, 1124-1132 (1985)

[35] What Are Log-Gabor Filters and Why Are They Good? - http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/PhaseCongruency/Docs/convexpl.html Viewed in December 2008

[36] D. Field, "Relations between the statistics of natural images and the response properties of cortical cells," Journal of Optical Society of America, vol. 4, no. 12, pp. 2379–2394, 1987

[37] Makrogiannis, S.K. Bourbakis, N.G. "A stochastic optimization scheme for automatic registration of aerial images", Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on, 15-17 Nov. 2004. pp. 328- 336.

[38] Jiangjian Xiao, Yunjun Zhang, Mubarak Shah, "Adaptive Region-Based Video Registration," wacv-motion, pp. 215-220, IEEE Workshop on Motion and Video Computing (WACV/MOTION'05) - Volume 2, 2005.

[39] A. Karargyris and N. Bourbakis, A methodology for reducing WCE video frames by using image registration, ATRC TR-10, 2007

[40] S. Tsevas, D. Iakovidis, D. Maroulis, E. Pavlakis; Automatic Frame Reduction of Wireless Capsule Endoscopy Video, 8th IEEE International Conference on BioInformatics and BioEngineering

[41] Median as a weighted arithmetic mean of all sample observations, SK Mishra, Economics Bulletin Journal, http://economicsbulletin.vanderbilt.edu/2004/volume3/EB-04C10011A.pdf - Viewed in November 2008

[42] S. Liangpunsakul, L. Mays and DK Rex. "Performance of Given suspected blood indicator" American Journal of Gastroenterology, Vol. 89. No. 12, 2003

[43] Does the "Suspected Blood Indicator" improve the detection of bleeding lesions by capsule endoscopy? Gastrointestinal Endoscopy, Volume 61, Issue 2, Pages 243-249, P.D'Halluin, M.Delvaux, M.Lapalus, S.Sacher-Huvelin, E.Soussan, L.Heyries, B.Filoche, J.Saurin, G.Gay, D.Heresbach

[44] Ramirez F. Gastroenterology and Hepatology: Small Intestine. Ed. by Mark Feldman (series editor), Lawrence R. Schiller. ©1997 Current Medicine Group LLC.

[45] Pennazio M, Santucci R, Rondonotti E, et al. Wireless capsule endoscopy in patients with obscure gastrointestinal bleeding: Results of the Italian multicenter experience. Gastrointestinal En3doscopy 2002; 55:AB87

[46] Hara AK, Leighton JA, Sharma VK, Fleischer DE. Small bowel: preliminary comparison of capsule endoscopy with barium study and CT. Radiology 2004; 230:260–265

[47] Radiological Society of North America, Radiographics Journal, Imaging of Small Bowel Disease: Comparison of Capsule Endoscopy, Standard Endoscopy, Barium Examination, and CT, http://radiographics.rsnajnls.org/cgi/content/full/25/3/697#F3A , Article viewed in May 2008.

[48] S. Hwang, J. Oh, J. Cox, S. J. Tang, and H. F. Tibbals. Blood detection in wireless capsule endoscopy using expectation maximization clustering, volume 6144. SPIE, 2006.

[49] P.Y. Lau, P. L. Correia, "Detection of Bleeding Patterns in WCE Video using Multiple Features", Proceedings of the Int. Conf. of the IEEE Engineering in Medicine and Biology Society, EMBC'2007, Lyon, France, Aug. 2007

[50] Y. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation," Computer Graphics and Image Processing, Vol. 13, No. 3, July, 1980, pp. 222 - 241.

[51] Applied and Computational Control,. Signals, and Circuits—Volume 1, edited by Biswa Nath Datta, Birkhäuser,. 1999, 539 pp., ISBN 0-8176-3954-3

[52] J. Kender, Saturation, Hue and Normalized Color: Calculation, Digitization Effects and Use, Technical Report, Dept of CS, Carnegie-Mellon University, 1976.

[53] What I need to know about Colon Polyps? National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health. http://digestive.niddk.nih.gov/ddiseases/pubs/colonpolyps_ez/ - Viewed in November 2008.

[54] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, "Computer Aided Tumor Detection in Endoscopic Video using Color Wavelet Features", IEEE Transactions on Information Technology in Biomedicine, vol. 7, 2003, pp. 141-152.

[55] Luís A. Alexandre, Nuno Nobre, João Casteleiro, "Color and Position versus Texture Features for Endoscopic Polyp Detection". BMEI (2) 2008, pp. 38-42.

[56] Dongqing Chen; Farag, A.A.; Hassouna, M.S.; Falk, R.; Dryden, G.W., "Geometric Features Based Framework for Colonic Polyp Detection using a New Color Coding Scheme," *Image Processing, 2007. ICIP 2007. IEEE International Conference on* , vol.5, no., pp.V -17-V -20, Sept. 16 2007-Oct. 19 2007

[57] Jianhua Yao; Miller, M.; Franaszek, M.; Summers, R.M., "Colonic polyp segmentation in CT colonography-based on fuzzy clustering and deformable models," *Medical Imaging, IEEE Transactions on* , vol.23, no.11, pp.1344-1352, Nov. 2004

[58] Konukoglu, E.; Acar, B.; Paik, D.S.; Beaulieu, C.F.; Rosenberg, J.; Napel, S., "Polyp Enhancing Level Set Evolution of Colon Wall: Method and Pilot Study," *Medical Imaging, IEEE Transactions on* , vol.26, no.12, pp.1649-1656, Dec. 2007

[59] Anna K. Jerebko, Sheldon Teerlink, Marek Franaszek, and Ronald M. Summers, "Polyp segmentation method for CT colonography computer-aided detection" ,Proc. SPIE 5031, 359 (2003), DOI:10.1117/12.480696

[60] Well-Connected Reports, Harvard Medical School, Peptic Ulcers, September 2001

[61] eMedicineHealth – Practical Guide to Health , http://www.emedicinehealth.com/peptic_ulcers/article_em.htm#Peptic%20Ulcers%20Overview (viewed in January 2009)

[62] Umaprasanna S. Karnam, Charles M. Rosen, Jeffrey B. Raskin - Small bowel ulcers, Current Treatment Options in Gastroenterology Journal, Vol. 4, Issue 1, pp 15-21

[63] P.Y. Lau, P. L. Correia, "Detection of Bleeding Patterns in WCE Video using Multiple Features", Proceedings o f the Int. Conf of the IEEE Engineering in Medicine and Biology Society, EMBC' 2007, Lyon, France, Aug. 2007

[64] A Survey on Wireless Capsule Endoscopy and Endoscopic Imaging, Engineering in Medicine and Biology Society Journal (to appear)

[65] Karargyris, Alexandros; Bourbakis, Nikolaos, "A methodology for detecting blood-based abnormalities in Wireless Capsule Endoscopy videos," BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on , vol., no., pp.1-6, 8-10 Oct. 2008

[66] Baopu Li,; Meng, Max Q.-H., "Ulcer recognition in capsule endoscopy images by texture features," Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on , vol., no., pp.234-239, 25-27 June 2008

[67] Moghaddamzadeh, A. and Bourbakis, N., "A fuzzy region growing approach for segmentation of color images". Pattern Recognition. v30 i6. 867-881.

[68] N. Bourbakis," Detecting abnormal patterns in WCE images", IEEE Symp. BIBE-05, Oct. 17-19, 2005, MN, pp. 232-238.

[69] J. Daugman. Uncertainty relation for resolution in space, spatial frequency and orientation optimized by twodimensional visual cortical filters. J. Opt. Soc. Am., 2(7):1160–1169, 1985.

[70] M. A. Webster and R. L. De Valois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," J. Opt. Soc. Am. A **2**, 1124-1132 (1985)

[71] What Are Log-Gabor Filters and Why Are They Good? - http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/PhaseCongruency/Docs/convexpl.html Viewed in December 2008

[72] D. Field, "Relations between the statistics of natural images and the response properties of cortical cells," Journal of Optical Society of America, vol. 4, no. 12, pp. 2379–2394, 1987.

[73] S.M. Smith. "SUSAN - a new approach to low level image processing" Internal Technical Report TR95SMS1, Defence Research Agency, Chobham Lane, Chertsey, Surrey, UK, 1995

[74] Chan, T.F.; Vese, L.A., "Active contours without edges," Image Processing, IEEE Transactions on , vol.10, no.2, pp.266-277, Feb 2001

[75] Kodogiannis, Vassilis and Boulougoura, Maria (2007) An adaptive neurofuzzy approach for the diagnosis in wireless capsule endoscopy imaging. International Journal of Information Technology, 13 (1). pp. 46-56. ISSN 1305-2403

[76] Tkalcic, M.; Tasic, J.F., "Colour spaces: perceptual, historical and applicational background," EUROCON 2003. Computer as a Tool. The IEEE Region 8 , vol.1, no., pp. 304-308 vol.1, 22-24 Sept. 2003

[77] Y. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation," Computer Graphics and Image Processing, Vol. 13, No. 3, July, 1980, pp. 222 – 241

[78] Vapnik, Vladimir N (1989). 'Statistical Learning Theory'. Wiley-Interscience. ISBN 0-471-03003-1.

[79] Haralick, R. M., Shanmugam, K., Dinstein, I., "Textural Features for Image Classification", IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC 3, No. 6, Nov 1973

[80] Inoue, T.; Abe, S., "Fuzzy support vector machines for pattern classification," Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on , vol.2, no., pp.1449-1454 vol.2, 2001

[81] Umaprasanna S. Karnam, Charles M. Rosen, Jeffrey B. Raskin - Small bowel ulcers, Current Treatment Options in Gastroenterology Journal, Vol. 4, Issue 1, pp 15-21

[82] Weka 3: Data Mining Software, University of Waikato , http://www.cs.waikato.ac.nz/ml/weka/

[83] Karargyris, A.; Bourbakis, N., "Identification of ulcers in Wireless Capsule Endoscopy videos," *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on* , vol., no., pp.554-557, June 28 2009-July 1 2009

[84] U. Mueller-Richter, A. Limberger, P. Weber, K. Ruprecht, W. Spitzer, and M. Schilling. Possibilities and limitations of current stereo-endoscopy. In Surgical Endoscopy, Volume 18, Number 6, pages 942-947, June 2004.

[85] James R. Zekta Jr., Surgeons and the Scope, Cornell University Press, 2003.

[86] F. Mourgues; F. Devemay and E. Coste-Maniere. 3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery, In Proceedings. IEEE and ACM International Symposium on Augmented Reality

[87] T. Thormahlen, H Broszio, and P. Meier. Three dimensional endoscopy. In Falk Symposium, No. 124, 2002.

[88] C. Wu, Y. Chen, C. Liu, C. Chang, and Y. Sun. Automatic extraction and visualization of human inner structures from endoscopic image sequences. In Proceedings of the SPIE, Volume 5369, pages 464-473, 2004.

[89] Q. Liu, R.J. Sclabassi, N. Yao, and M. Sun. 3D construction of endoscopic images based on computational stereo. In Bioengineering Conference, pages 69-70, 2006

[90] C. Wu, Y. Sun, and C. Chang. Three-Dimensional Modeling from Endoscopic Video Using Geometric Constraints Via Feature Positioning. In IEEE Transactions on Biomedical Engineering. Volume 54, No. 7, July 2007

[91] Jin Zhou; Das, A.; Feng Li; Baoxin Li, "Circular generalized cylinder fitting for 3D reconstruction in endoscopic imaging based on MRF," Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on , vol., no., pp.1-8, 23-28 June 2008

[92] R.I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge Univ. Press, 2003.

[93] M. Pollefeys, L. J. V. Gool, M. Vergauwen, F. Verbiest, K.Cornelis, J. Tops, R. Koch. Visual Modeling with a Hand-Held Camera, In International Journal of Computer Vision, Volume 59, Issue 3, pages 207-232, 2004.

[94] 3D Surface Reconstruction from Endoscopic Videos

[95] J.P. Helferty and W.E. Higgins. Combined endoscopic video tracking and virtual 3D CT registration for surgical guidance. In ICIP, Volume 2, pages 961-964, 2002.

[96] D. Burschka, M. Li, M. Ishii, R. H. Tylor and G. D. Hager. Scale-invariant registration of monocular endoscopic images to CT-Scans for sinus surgery. In Medical Image Analysis. Volume 9, Issue 5, pages 413-426, October 2005.

[97] H. Mitsuhiro, S. Naoki and N. Yoshihiko. Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management. In Medical Image Analysis, volume 10, Issue 4, pages 509-519, August 2006.

[98] D. Tomazevic, B. Likar, and F. Pernus. 3-D/2-D registration by integrating 2-D information in 3-D. In IEEE Transactions on Biomedical Engineering. Volume 25, No. 1, January 2006

[99] A system for an accurate 3D reconstruction in. Video Endoscopy Capsule. A. Kolar, O. Romain, J. Ayoub. D. Faura, S. Viateur. Universite P&M CURIE - Paris VI

[100] P. M. Szczypinski, P. V. J. Sririam, R. D. Sririam, and D. Reddy, "Model of deformable rings for aiding the wireless capsule endoscopy video interpretation and reporting," in Proc. Int. Conf. Computer Vision Graphics, Warsaw, Poland, 2004, pp. 22–24.

[101] Ruo Zhang , Ping-Sing Tsai , James Edwin Cryer , Mubarak Shah, Shape from Shading: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.21 n.8, p.690-706, August 1999

[102] B. K. P. Horn. Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View. PhD thesis, MIT, 1970

[103] "A rigorous and realistic Shape from Shading method and some of its applications" (Emmanuel Prados and Olivier Faugeras). INRIA Research Report No RR-5133, March 2004

[104] J. Oliensis. Shape from shading as a partially well-constrained problem. Computer Vision, Graphics, and Image Processing: Image Understanding, 54:163{183, 1991

[105] P.S. Tsai and M. Shah. Shape from shading using linear approximation. Image and Vision Computing Journal, 12(8):487{498, 1994

[106] J.H. Lambert, Photometria Sive de Mensura de Gratibus Luminis, Colorum et Umbrae. Augsberg, Germany: Eberhard Klett, 1760.

[107] Tan, R.T.; Ikeuchi, K., "Separating reflection components of textured surfaces using a single image," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.27, no.2, pp.178-193, Feb. 2005

[108] B. Horn and B. Schunck. Determining optical flow. Artificial Intelligence, 17:185–203, 1981.

[109] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," ACM Comput. Surveys, vol. 27, no. 3, pp. 444–467, Sept. 1995.

[110] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In IEEE Int. Conf. on Computer Vision, pages 231–236, 1993.

[111] H. Nagel andW. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. IEEE Trans. on Patt. Anal. andMach.Intell., 8(5):565–593, 1986.

[112] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In ECCV, volume 3024, pages 25–36, 2004.

[113] L. Alvarez, J.Weickert, and J. S´anchez. Reliable estimation of dense optical flow fields with large displacements. International Journal of Computer Vision, 39(1):41–56, Aug. 2000.

[114] E.Memin and P. Perez. Hierarchical estimation and segmentation of dense motion fields. Int. J. of Computer Vision, 46(2):129–155, 2002

[115] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In IEEE Int. Conf. on Computer Vision, 2007.

[116] F. Steinbr ¨ucker, T. Pock, and D. Cremers. Large displacement optical flow computation without warping. In IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan, 2009.

[117] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. In Proc. IEEE Conf. on Comp. Vision Patt. Recog.(CVPR'09), 2009.

[118] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Anal. Mach. Intell., 25(2):218–233, 2003.

[119] 3D representation of the digestive tract surface in Wireless Capsule Endoscopy videos, Alexandros Karargyris, Orestis, Karargyris, Dr. Nikolaos Bourbakis, Technical Report November 2009

[120] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A Multillinear Singular Value Decomposition," SIAM J. Matrix Anal. Appl., vol. 21, no. 4, pp. 1253–1278, 2000.

[121] Nyugen, L.; Xu, Y.; Roy-Chowdhury, A., "An Illumination Invariant 3D Model Based Tracking Algorithm, with Application in Video Compression," Image Processing, 2006 IEEE International Conference on , vol., no., pp.1213-1216, 8-11 Oct. 2006

[122] A Model of Brightness Variations Due to Illumination Changes and Non-rigid Motion Using Spherical Harmonics, J. M. Buenaposada, Alessio Del Bue, E. Muñoz, L. Baumela, Proc of BMVC 2008 - 19th British Machine Vision Conference, University of Leeds, Leeds, UK, 2008

[123] Ferrant, M., Warfield, S. K., Guttmann, C. R. G., Mulkern, R. V., Jolesz, F. A., and Kikinis, R. 1999. 3D matching using a finite element based elastic deformation model. In *Proc. MICCAI'98,* pp. 202–209

[124] "A technique for blood detection in Wireless Capsule Endoscopy images" Barbara Penna (Politecnico di Torino, Italy); Tammam Tillo (Xi'an Jiaotong-Liverpool University, P.R. China); Marco Grangetto (Universita' di Torino, Italy); Enrico Magli (Politecnico di Torino, Italy); Gabriella Olmo (Politecnico di Torino, Italy) pp. 1864-1868, 17th European Signal Processing Conference (EUSIPCO 2009) Glasgow, Scotland, August 24-28, 2009

[125] I.S. Reed and X. Yu, "Adaptive multiband cfar detection of an optical pattern with unknown spectral distribution," in IEEE Trans. on Acoustics, Speech and Signal Processing, Oct. 1990, vol. 38/10, pp. 1760 – 1770.

[126] Yun Sub Jung; Young Ho Kim; Dong Ha Lee; Jong Hyo Kim; , "Active Blood Detection in a High Resolution Capsule Endoscopy using Color Spectrum Transformation," *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on* , vol.1, no., pp.859-862, 27-30 May 2008

[127] Giritharan, Balathasan; Xiaohui Yuan,; Jianguo Liu,; Buckles, Bill; JungHwan Oh,; Shou Jiang Tang,; , "Bleeding detection from capsule endoscopy videos," *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* , vol., no., pp.4780-4783, 20-25 Aug. 2008

[128] Baopu Li; Meng, M.Q.-H.; , "Computer-Aided Detection of Bleeding Regions for Capsule Endoscopy Images," *Biomedical Engineering, IEEE Transactions on* , vol.56, no.4, pp.1032-1039, April 2009

[129] Barbosa, D.J.C.; Ramos, J.; Correia, J.H.; Lima, C.S.; , "Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors," *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE* , vol., no., pp.6683-6686, 3-6 Sept. 2009

[130] Baopu Li; Meng, M.Q.-H.; , "Ulcer recognition in capsule endoscopy images by texture features," *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on* , vol., no., pp.234-239, 25-27 June 2008

[131] N.Bourbakis, et. al. Object recognition using Local – Global Graphs, Int. IEEE Conf. on TAI-03, pp. 616- 627,CA, Nov. 2003

[132] Zvi Fireman and Yael Kopelman, "New frontiers in capsule endoscopy", Journal of Gastroenterology and Hepatology: Review, 22 (2007) 1174–1177

[133] Yawen Chan, Max Q.-H. Meng and Xiaona Wang, A Prototype Design of Wireless Capsule Endoscope, Proc. of the IEEE International Conference on Mechatronics & Automation, Niagara Falls, Canada • July 2005

[134] A. Menciassi, M. Quirini, & P. Dario, Microrobotics for future gastrointestinal endoscopy, Minimally Invasive Therapy. 2007; 16:2; 91–100

[135] P. Valdastri, C. Quaglia E. Susilo, A. Menciassi, P. Dario, C. N. Ho, G. Anhoeck, M. O. Schurr, Wireless therapeutic endoscopic capsule:in vivo experiment, Endoscopy 2008; 40: 979-982

[136] Glass, P.; Cheung, E.; Sitti, M.; , "A Legged Anchoring Mechanism for Capsule Endoscopes Using Micropatterned Adhesives," Biomedical Engineering, IEEE Transactions on , vol.55, no.12, pp.2759-2767, Dec. 2008

[137] R. N. Miftahof, "The wave phenomena in smooth muscle syncytia," In Silico Biol., vol. 5, no. 5–6, pp. 479–498, Nov. 2005.

[138] N. Bourbakis, A. Karargyris and G. Giakos, Design a New Generation of Robotic Capsules in Endoscopy, IEEE Conf on Imaging Technologies and Systems, July 1-3, 2010, Greece.

ALEXANDROS KARARGYRIS Curriculum Vitae

Alexandros Karargyris graduated from National Technological University of Athens, Greece with a Masters degree in Electrical Engineering and Computer Engineering in 2006. He graduated from Wright State University, Dayton, OH, USA with a Ph,D. in Computer Engineering in 2010. His current research area is the development of computer aided detection algorithms for improving the usability and reducing process time of Wireless Capsule Endoscopy videos. He is interested in computer vision and medical imaging analysis. He is a holder of NSF and Ohio Board of Regents (OBR) scholarships. Currently he is a post-doctorate fellow at the National Library of Medicine (NLM) of National Institutes of Health (NIH).